

# **TAMING KILLER ROBOTS** Giving meaning to the "Meaningful Human Control" Standard for Lethal Autonomous weapon systems

Adam Cook Lieutenant Colonel, USAF

**The JAG School Papers** 

150 Chennault Circle, Building 694 Maxwell AFB, AL 36112

## JUDGE ADVOCATE GENERAL'S SCHOOL

## UNITED STATES AIR FORCE



# **Taming Killer Robots**

## Giving Meaning to the "Meaningful Human Control" Standard for Lethal Autonomous Weapon Systems

Adam Cook, Lieutenant Colonel, USAF

JAG School Paper no. 1

Air University Press Curtis E. LeMay Center for Doctrine Development and Education Maxwell Air Force Base, Alabama Project Editor Donna Budjenska

*Editorial Assistant* Tammi K. Dacus

*Cover Art, Book Design, and Illustrations* L. Susan Fair

*Composition and Production* Megan N. Hoehn

#### AIR UNIVERSITY PRESS

Director, Air University Press Lt Col Darin Gregg

Air University Press 600 Chennault Circle, Building 1405 Maxwell AFB, AL 36112-6010 https://www.airuniversity.af.edu/AUPress/

Facebook: https://www.facebook.com/AirUnivPress

and

Twitter: https://twitter.com/aupress



Accepted by Air University Press in October 2017 and published June 2019. ISSN: 2643-8933

#### Disclaimer

Lieutenant Colonel Adam Cook serves in the United States Air Force (USAF) Judge Advocate General's Corps. Opinions, conclusions, and recommendations expressed or implied within are solely those of the author and do not necessarily represent the views of the Judge Advocate General's Corps, the Air University, the USAF, the Department of Defense, or any other US government agency. Cleared for public release: distribution unlimited.

This JAG Paper and others in the series are available electronically at the Air University Press website: https://www.airuniversity .af.edu/AUPress/.



### Contents

The JAG School Papers	iv
Introduction	1
Meaning of Autonomy	2
Lethal Autonomous Weapon Systems Terminology	3
Autonomous Weapon Systems and the Laws of	
Armed Conflict	6
The Case for a Preemptive Ban	7
Pushback to Human Rights Watch's Views	9
A Third Perspective	12
Meaningful Human Control	16
Conclusion	23
Bibliography	24

#### **The JAG School Papers**

The JAG School Papers and other scholarly works published by Air University Press (AUP) provide independent analysis and constructive discussion on issues important to Air Force commanders, staffs, and other decision makers. Each paper can also be a valuable tool for defining further research. To make comments about this paper or submit a manuscript to be considered for publication in this series, please email The Air Force Judge Advocate General's School at afloa.afjags@us.af.mil. For other queries or submissions, please email AUP at lemaycenter.au.press@us.af.mil.

The JAG School Papers showcase top submissions received for the annual writing competition hosted by the Air Force Judge Advocate General's School in partnership with the Air Force JAG School Foundation Inc. The competition provides a forum for students, practitioners, academics, and policy makers to provide unique, original, previously unpublished perspectives and insights.

#### Introduction

From Leonardo da Vinci's mechanical knight to the malevolent Skynet from the fictional *Terminator* series, humans the world over have been simultaneously fascinated and terrified by the potential of robotics and artificial intelligence to fundamentally change, or potentially end, human life on this planet. All of these hopes and fears, accumulated during the centuries, have come to a head as technologies have caught the tail of fantasy and systems once found only in dreams, or nightmares, which have moved to the precipice of reality.

This robotics revolution, represented by new systems ranging from the automated Roomba vacuum cleaner to guided missiles and automated interceptors, has understandably led to a heated debate over the rules of the road in this new era. This discussion is particularly contentious in the area of weapon systems (whether real or imagined) guided in large part by artificial intelligence. Do these systems just represent the latest iteration in the evolution of weapons dating to the dawn of human existence, or do they represent something fundamentally different than all the systems to come before them? Are the traditional laws of armed conflict, as detailed in The Hague and Geneva Conventions, sufficient to protect human rights in the robotic age? What, if anything, do these century-old principles have to say about autonomous weapon systems?

These questions and others have led divergent groups of scholars, institutions, and government officials to stake out seemingly irreconcilable positions on the principles of warfare in the robotics age. These views range from calls for an outright, and preemptive, ban on all lethal autonomous weapon systems to the view that the current principles of lawful warfare are enduring and that continued application of these rules to new systems on a case-by-case basis is the best way to protect civilians and minimize unnecessary suffering.

This paper will assess these disparate views and ultimately stake out a third position: while autonomous weapon systems do not, by their nature, violate the current laws of war, they do represent a fundamental departure from all the weapons to come before them. As such, this paper recommends the adoption of an additional standard: a requirement of "meaningful human control" over all weapons systems developed or deployed by a nation-state. As detailed in this paper, however, despite its popularity in many corners of the international community, the meaningful human control standard is useless, and potentially harmful, without further refinement of what such a standard means in practice. As such, this paper recommends adoption of a three-factor test, including an assessment of (1) time, (2) geography, and (3) discernment

in determining whether any proposed system is compliant with the meaningful human control standard. It is my hope that the proposal of such a standard will further spur debate over the practical application of any proposed regulation of autonomous weapon systems at the national or international level.

#### **Meaning of Autonomy**

Any useful discussion of lethal autonomous weapon systems (LAWS)<sup>1</sup> must begin with an understanding of what it means for a system to be autonomous. As will be seen, much of the disagreement over the legality and advisability of developing and utilizing LAWS is driven by an (often unstated) disagreement over what exactly is meant by an autonomous or fully autonomous weapon system. As Paul Scharre and Michael Horowitz explain in their ethical autonomy working paper, "There is no internationally agreed-upon definition of what constitutes an 'autonomous weapon,' making clear communication on the topic more difficult. . . . This lack of clarity in terminology is further compounded by the fact that some are calling for autonomous weapons to be regulated or banned even before consensus exists on how to define the category. Thus, at present, definitions are tied up in debates over the technology itself. . . . [This] lack of clarity on basic terminology itself is a recipe for disaster."<sup>2</sup>

Despite these semantic disagreements, there is a general consensus that an autonomous system is one that can execute complex tasks over a wide range of circumstances without direct human involvement. As Scharre and Horowitz write, "In its simplest form, autonomy is the ability of a machine to perform a task without human input. Thus, an 'autonomous system' is a machine, whether hardware or software, that, once activated, performs some task or function on its own."<sup>3</sup>

Rather than a fundamentally distinct category, however, autonomy is better thought of as existing on the far end of a continuum of increasingly sophisticated systems ranging from purely manual machines, to simple automatic systems, to automated devices, and finally autonomous systems. As Christof Heyns, the United Nations special rapporteur on extrajudicial, summary or arbitrary executions, wrote in a recent report on LAWS, "Autonomous' needs to be distinguished from 'automatic' or 'automated.' Automatic systems, such as household appliances, operate within a structured and pre-

<sup>1.</sup> Throughout this paper, the terms Lethal Autonomous Weapon systems (LAWS), autonomous weapon systems (AWS), and fully autonomous weapons (FAW) are generally used interchangeably, unless a specific meaning is detailed.

<sup>2.</sup> Scharre and Horowitz, "Introduction to Autonomy," 3.

<sup>3.</sup> Scharre and Horowitz, 5.

dictable environment. Autonomous systems can function in an open environment, under unstructured and dynamic circumstances."<sup>4</sup>

The Air Force's recent publication *Autonomous Horizons: System Autonomy in the Air Force—A Path to the Future*, vol. 1: *Human-Autonomy Teaming*, further explains:

In general [an autonomous system] involves the use of additional sensors and more complex software to provide higher levels of automated behaviors over a broader range of operating conditions and environmental factors, and over a wider range of functions or activities. Autonomy is often characterized in terms of the degree to which the system has the capability to achieve mission goals independently, performing well under significant uncertainties, for extended periods of time, with limited or non-existent communication, and with the ability to compensate for system failures, all without external intervention. . . Autonomy can be considered as well-designed and highly capable automation.<sup>5</sup>

As an example to illustrate this progression, in many of our lifetimes we have seen the basic passenger car evolve from a strictly manual device with a mechanical gearshift, to one with an automatic transmission which independently shifts gears based on driver inputs, to the use of increasingly sophisticated automated systems and sensors which assist the driver in tasks including navigating, maintaining a constant speed, parking, and even changing lanes. In the near future, we may see fleets of autonomous vehicles which are networked to our personal calendars and will pick us up at our homes at the appropriate time, select the best route to our scheduled destination, navigate, and drop us off—all with no direct human intervention or control. While it will be difficult to identify the exact moment in its evolution when a passenger car becomes truly "autonomous," it is clear that driverless cars present fundamentally different opportunities, risks, and policy issues than Henry Ford's iconic Model T. The same is true in the emerging field of autonomous weapon systems.

### Lethal Autonomous Weapon Systems Terminology

Beyond the varied attempts to define autonomy in general, a distinct vernacular has developed in recent years around the debate over autonomous weapon systems. In the international community, academia, and the world's militaries, terms such as fully autonomous, semi-autonomous, human (in, on, out) of the loop, and meaningful human control have been frequently used to frame key aspects of the ongoing debate.

<sup>4.</sup> Heyns, Report of the Special Rapporteur, 8.

<sup>5.</sup> See USAF Office of the Chief Scientist, Autonomous Horizons, vol. 1, 3-4.

Before touching on each of these terms briefly, it is important to understand that, at its core, the controversy over the legality and humanity of autonomous weapon systems is focused almost entirely on the question of target selection. As their various writings have made clear, the principal objection raised by Human Rights Watch (HRW) and other skeptics to the development of LAWS is the possibility that an autonomous system (or killer robot, in HRW's more evocative formulation) would have the ability to independently select and engage targets of its own volition, without direct authorization by a human. Each of the key terms discussed below is used to further explore this fundamental concern.

The *human in the loop, human on the loop*, and *human out of the loop* formulation was developed in the early days of the debate over LAWS to distinguish between the degrees of automation inherent in various existing, planned, and imagined weapon systems.<sup>6</sup> A human in the loop system is one in which a human being selects a target and then the weapon or weapon system uses some combination of sensors, navigational aids, and automated processes to track and engage the target. Human in the loop systems are neither new nor particularly controversial and have been utilized by the world's militaries since the first homing torpedoes employed during World War II.<sup>7</sup> Examples of modern human in the loop systems range from the full catalog of laser or GPS-guided munitions to homing devices such as guided air-to-air missiles.

A human on the loop weapon system, by contrast, is one which can select and engage targets independently but operates under the direct supervision of a human operator with the ability to intervene and cancel engagements when necessary. These systems, which utilize preset parameters in combination with inputs from their sensors to identify authorized targets, have to date been used exclusively in defensive systems used to protect installations, ships, or human populations from incoming missiles, mortars, or artillery barrages. Examples include Israel's Iron Dome antimissile system, the US Navy's Phalanx *Aegis*-class ship defense system, and the US Army's Patriot batteries.

Finally, a human out of the loop system is one which operates in at least some circumstances truly independently. That is, it can select and engage targets without any contemporaneous oversight or possibility of intervention by a human operator (though it would, presumably, be acting in accordance with parameters input by a human prior to deployment). The only examples of truly human out of the loop weapon systems which have been deployed—or even developed to date—are loitering munitions, which are programmed to

<sup>6.</sup> Scharre and Horowitz, "An Introduction to Autonomy," 8-15; and Human Rights Watch, Losing Humanity, 2.

<sup>7.</sup> Scharre and Horowitz, "An Introduction to Autonomy," 8.

search a defined area for specified types of targets (such as ships, tanks, or radars) and then fly into and destroy targets meeting the given specifications. While the US Navy has developed experimental versions of these types of weapons, the only currently operational loitering munition appears to be Israel's Harpy, an antiradar weapon that is designed to search for and destroy radar systems meeting its preset criteria within a given target area.<sup>8</sup> As Scharre details in his follow-on paper on ethical autonomy, "Autonomous Weapons and Operational Risk," the Harpy has the ability to stay aloft and search for targets for up to two-and-one-half hours, during which time it is out of the control of any human operator.<sup>9</sup>

Along with, and sometimes dependent upon, the on/in/out of the loop nomenclature, many of the analyses of LAWS use the terms semi-autonomous versus fully autonomous or simply autonomous to draw key distinctions. The HRW, in its eport *Losing Humanity: The Case against Killer Robots*, explicitly relies on the in/on/out of the loop distinction in its definition of a fully autonomous system, explaining that, in its view, "The term 'fully autonomous weapon' refers to both out-of-the-loop weapons and those that allow a human on the loop, but that are effectively out-of-the-loop weapons because the supervision is so limited."<sup>10</sup>

In its official policy statement on autonomous weapon systems, the Department of Defense (DOD) avoids the in/on/out of the loop terminology, but defines an autonomous weapons system as one that "once activated, can select and engage targets without further intervention by a human operator."<sup>11</sup> Along similar lines, Professor Michael Schmitt and Army Lt Col Jeffrey Thurnher write that "the crux of full autonomy... is the capability to identify, target and attack a person or object without further human input after activation."<sup>12</sup>

In the DOD formulation, a fully autonomous or simply autonomous system is functionally and legally distinct from a semi-autonomous system, which DODD 3000.09 defines as "a weapon system that, once activated, is intended to only engage specific targets or specific target groups that have been selected by a human operator." The Directive specifically mentions fireand-forget homing munitions such as guided air-to-air missiles as examples of semi-autonomous weapon systems.<sup>13</sup> Without using the term, the DOD's

<sup>8.</sup> Scharre and Horowitz, "An Introduction to Autonomy," 13–14.

<sup>9.</sup> Scharre, "Autonomous Weapons and Operational Risk," 21.

<sup>10.</sup> Human Rights Watch, Losing Humanity, 2.

<sup>11.</sup> Department of Defense Directive (DODD) 3000.09, Autonomy in Weapon Systems, 14.

<sup>12.</sup> Schmitt and Thurnher, "Out of the Loop," 235.

<sup>13.</sup> DODD 3009.09, 14.

definition of a semi-autonomous weapon is thus functionally identical to the commonly accepted understanding of a human in the loop system.

To further confuse the debate, DODD 3000.09 also uses the term "humansupervised autonomous weapon system," which is defined as "an autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur."<sup>14</sup> As the discerning reader will note, this is essentially identical to the commonly understood definition of a human on the loop system.

While the above definitions in some ways raise more questions than they answer, they do at least lay a foundation for identifying the bases of disagreement between the various parties to the debate over LAWS.

## Autonomous Weapon Systems and the Laws of Armed Conflict

As with any new, or potentially new, type of weapon systems, nations seeking to develop and deploy LAWS are required by international law to perform a legal analysis of the proposed system's compatibility with the international laws concerning armed conflict, including the Geneva and Hague conventions. Article 36 of Additional Protocol 1 to the Geneva Conventions states that "in the study, development or adoption of a new weapon, means or method of war, a High Contracting Party<sup>15</sup> is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."<sup>16</sup> While there is wide agreement that such a legal analysis is required by all nations seeking to develop and deploy LAWS, there is vigorous disagreement over the ultimate question of whether such weapons inherently violate the international laws of armed conflict.

As the debate over the legality of LAWS has evolved, three distinct legal positions have developed. The first, argued most visibly and passionately by HRW, is that LAWS by their very nature do not, and cannot, comply with international law. The second position, explicated most clearly by Schmitt and Thurnher, is that LAWS simply represent the latest generation of technological change in warfare and while, like any other weapon systems, they can

<sup>14.</sup> DODD 3009.09, 14.

<sup>15.</sup> While the US is not a signatory to Additional Protocol 1 to the Geneva Convention, the requirement to conduct legal reviews of new weapon systems is widely considered to be customary international law and therefore binding on all nations. *See* Schmitt and Thurnher, "Out of the Loop," 271.

<sup>16.</sup> Protocol Additional to the Geneva Conventions of 12 August 1949, Article 36.

certainly be *used* in ways that violate international law, there is nothing about LAWS that represents a *per se* violation of the laws of war. The final position, and the one taken by this paper, is that while LAWS do not by their nature represent a violation of existing principles of the Law of Armed Conflict, they do represent a fundamentally new era in warfare which will require additional laws and standards to address the unique risks represented by self-targeting systems. Each of these arguments will be discussed.

#### The Case for a Preemptive Ban

In their provocative and high-profile report, *Losing Humanity: The Case against Killer Robots*, HRW and the International Human Rights Clinic at Harvard Law School argue that any fully autonomous weapon systems represent a *per se* violation of international law and should therefore be banned by the international community. As the report states in its summary, "Fully autonomous weapon systems would not only be unable to meet legal standards but would also undermine essential non-legal safeguards for civilians. Our research and analysis strongly conclude that fully autonomous weapons should be banned and that governments should not pursue that end."<sup>17</sup>

In their analysis of the application of international law to autonomous weapon systems, HRW focuses heavily on the law of armed conflict's principles of distinction, proportionality, and military necessity. The rule of distinction, which HRW refers to as "arguably the bedrock of international humanitarian law,"<sup>18</sup> and which is codified at Article 48 of Additional Protocol 1 of the Geneva conventions, mandates that combatants must "distinguish between the civilian population and combatants." The conventions also ban military actions that "employ a method or means of combat which cannot be directed at a specific military objective" or "employ a method or means of combat the effects of which cannot be limited."<sup>19</sup>

While acknowledging the potential for increasingly sophisticated sensors and processors to aid in the decision-making capacity of autonomous weapons, HRW nonetheless argues that LAWS would be incapable of accurately distinguishing between combatants and protected civilians: "Fully autonomous weapons would not have the ability to sense or interpret the difference between soldiers and civilians, especially in contemporary combat environments." Focusing on the asymmetric conflicts of recent years in which combatants have intentionally blended in with civilian populations, often in

<sup>17.</sup> Human Rights Watch, Losing Humanity, 2.

<sup>18.</sup> Human Rights Watch, 24.

<sup>19.</sup> Protocol Additional to the Geneva Conventions of 12 August 1949, Article 45 and 51(4).

densely populated urban environments, HRW reasons that, "in these conflicts, combatants do not wear uniforms or insignia. Instead they seek to blend in with the civilian population and are frequently identified by their conduct, or their 'direct participation in hostilities.'" In HRW's view, autonomous weapon systems would be incapable of distinguishing between combatants and civilians in this type of environment for a number of reasons. First, a robot (using HRW's terminology) might not have adequate sensors. Second, and more significantly, in the absence of uniforms or insignia, the only way to distinguish combatants from civilians is through the attribution of intention, and "one way to determine intention is to understand an individual's emotional state, something that can only be done if the soldier has emotions."<sup>20</sup>

In addition to failing the distinction standard, HRW also argues that LAWS are inherently violative of international law based on their inability to perform the type of proportionality analysis required under the Geneva conventions. Such an analysis, while inherently subjective, is required to avoid attacks which "may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage expected."<sup>21</sup> In other words, to comply with international law, the military advantage of any proposed attack must be weighed against the expected loss of civilian life, civilian injuries, and damage to civilian property.

In HRW's assessment, compliance with the proportionality standard "requires human judgment that a fully autonomous weapon would not have."<sup>22</sup> Given the complexity and judgment required to evaluate the full context of any decision to launch an attack, HRW argues that "it is highly unlikely that a robot could be preprogrammed to handle the infinite number of scenarios it might face so it would have to interpret a situation in real time," potentially leading to "chaotic robot behavior with deadly consequences."<sup>23</sup> HRW's overall conclusion is that the proportionality test "requires more than a balancing of quantitative data, and a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality."<sup>24</sup> Even if a robot could be programmed to make such a judgment similarly to a reasonable human being, HRW argues, "It would fail to have other characteristics—such as the ability to understand humans and the

<sup>20.</sup> Human Rights Watch, Losing Humanity, 31.

<sup>21.</sup> Protocol Additional to the Geneva Conventions of 12 August 1949, Article 51(b).

<sup>22.</sup> Human Rights Watch, Losing Humanity, 32.

<sup>23.</sup> Human Rights Watch, 32.

<sup>24.</sup> Human Rights Watch, 33.

ability to show mercy—that are necessary to make wise legal and ethical choices beyond the proportionality test.<sup>25</sup>

Finally, HRW argues that fully autonomous weapon systems would also be incapable of determining whether any given action meets the standard of military necessity and that the deployment of fully autonomous systems into a battlefield containing civilians might violate the Martens clause of the Geneva conventions, which bans weapons which violate the "principles of humanity" or "the dictates of public conscience."<sup>26</sup> For all of these reasons, HRW concludes that autonomous weapon systems' "inability to meet the core principles of international humanitarian law would erode legal protections and lead fully autonomous weapons to endanger civilians during armed conflict."<sup>27</sup>

#### Pushback to Human Rights Watch's Views

Two teams of legal scholars, Kenneth Anderson, Daniel Reisner, and Matthew Waxman writing for the Naval War College, and Schmitt and Thurnher, writing in the Harvard National Security Law Journal, have taken the lead in refuting HRW's analysis and making the case that LAWS: (1) do not inherently violate international law, (2) may in fact lower the burden of war on civilians, and (3) should not be preemptively banned. In their 2014 analysis, Anderson, Reisner, and Waxman focus on the distinction between illegal weapons versus unlawful use of lawful weapons. The authors note that, under international law, weapons should only be declared inherently unlawful if the intended or primary use of the weapon would be unlawful. To justify a ban, therefore, it is not enough to show that a weapon *could* be used unlawfully, but rather that if used exactly as intended the weapon would be likely to violate principles of distinction, cause unnecessary suffering, or have effects which cannot be controlled (as is the case with certain nerve agents and other biological weapons). As the authors explain, "None of these rules renders a weapons system illegal per se solely on account of it being autonomous. If a fully autonomous weapon system were supplied with sufficiently reliable parameters and it were able to act on them so as to be able to strike specific targets on the same legal terms of discrimination that would apply to a human soldier, that the weapon system was 'autonomous' would not violate the 'indiscriminate by nature' rule."28

<sup>25.</sup> Human Rights Watch, Losing Humanity, 34.

<sup>26.</sup> Human Rights Watch, 35.

<sup>27.</sup> Human Rights Watch, 36.

<sup>28.</sup> Anderson, Reisner, and Waxman, "Adapting the Law," 400-401.

In countering HRW's argument that LAWS are simply incapable of applying international principles such as distinction in real-world environments, the authors concede that there may be circumstances, such as a crowded urban environment with combatants intermixed with civilians, in which present-day autonomous systems would have difficulty complying with the laws of war. However, rather than supporting HRW's conclusion, the authors conclude that "this is not to say that autonomous systems are thereby illegal. Quite the opposite, in some settings their use would be legal and in others illegal, depending on how technologies advance."<sup>29</sup>

After concluding that autonomy in and of itself does not make a weapon *per se* unlawful under existing principles of international law, Anderson, Reisner, and Waxman then analyze potential violations of the law of war in the *use* of lethal autonomous weapon systems. In assessing autonomous weapon systems' compliance with international law in real-world battlefield scenarios, the authors emphasize that, at least for the foreseeable future, commanders will be responsible for the conditions under which they deploy autonomous systems.

It is critically important to understand, that before an autonomous weapon system—like any weapon system—is used in a military operation, human commanders and those employing the weapon will generally continue to be expected to exercise judgment about the likely presence of civilians and the likelihood that they may be inadvertently harmed; expected military advantage... the weapon's capabilities, limitations and safety features; and many other factors. It is difficult to draw general conclusions in the abstract about the many complex legal issues involved in such scenarios.<sup>30</sup>

Like Anderson, Reisner, and Waxman, the writing team of Schmitt and Thurnher takes the position that LAWS are not, simply by virtue of being autonomous, inherently violative of the laws of war and that such systems can almost certainly be lawfully used in at least some environments, even at a relatively primitive level of development: "What has been missed in much of the dialogue so far is that even an autonomous weapon system that is completely incapable of distinguishing a civilian from a combatant or a military objective from a civilian object can be used lawfully in certain environments."<sup>31</sup> Schmitt and Thurnher also argue that such systems, once developed to an appropriate level of sophistication, may actually be *more* compliant with the laws of war than human combatants.

Instead of merely making choices in order to complete specific and defined tasks, general artificial intelligence systems will exhibit human-like cognitive abilities, enabling them to make decisions in response to complex problems and situations. The systems

<sup>29.</sup> Anderson, Reisner, and Waxman, "Adapting the Law," 402.

<sup>30.</sup> Anderson, Reisner, and Waxman, 405.

<sup>31.</sup> Schmitt and Thurnher, "Out of the Loop," 246.

will adapt and learn by observing their environment and their interaction with it. In fact, the President of the International Committee for the Red Cross . . . has posed the prospect of an autonomous system that might "be programmed to behave more ethically and far more cautiously on a battlefield than a human being."<sup>32</sup>

As Schmitt and Thurnher point out, the standard for autonomous weapon systems' compliance with the laws of war is not whether they are able to make perfect decisions at all times, but whether they are able to follow the principles of distinction, proportionality, and military necessity at least as well as human operators: "It must be emphasized that as a matter of law, more may not be asked of autonomous weapon systems than of human-operated systems."<sup>33</sup> The authors conclude that with increasingly sophisticated sensors tied to advanced artificial intelligence, in the near future autonomous systems may be able to distinguish between civilians and combatants at least as well as a human operator in at least some battlefield environments, and that it will be the responsibility of commanders to ensure that any autonomous weapon systems used are capable of distinguishing between civilians and combatants in the environment in which they are deployed.<sup>34</sup>

In assessing the capacity of LAWS to comply with the principles of proportionality and military necessity, Schmitt and Thurnher note the possibility of future development of artificial intelligence algorithms sufficiently powerful to perform a comprehensive proportionality analysis but concede that for the time being such determinations will continue to be made by commanders who determine when and where such systems will be deployed and the parameters they are given:

For the immediate future, though, the actual proportionality decision will continue to be made by humans—by deciding to launch the system into a particular environment, by deciding how to preprogram the system, or by revising the engagement criteria remotely. These humans remain fully responsible for compliance with the rule of proportionality . . . for example, the operator will have violated the rule if he or she approved use of an autonomous system with weapons that were insufficiently precise to be used in a particular setting . . . and, as a result, harm to civilians and civilian objects was reasonably likely to be excessive to the anticipated military gains.<sup>35</sup>

While arguing against a preemptive ban and emphasizing the potential advantages of autonomous weapon systems to both the militaries that employ them and noncombatant civilians in hostile areas, Schmitt and Thurnher nonetheless acknowledge the unique risks and policy issues inherent in the deployment and use of LAWS. The best way to combat misuse of autonomous

<sup>32.</sup> Schmitt and Thurnher, "Out of the Loop," 239-40.

<sup>33.</sup> Schmitt and Thurnher, 247.

<sup>34.</sup> Schmitt and Thurnher, 252-53, 278.

<sup>35.</sup> Schmitt and Thurnher, 257.

weapon systems and prevent the threats to civilian populations warned against by HRW, argue the authors, is to ensure that all future systems are legally reviewed both in their development and real-world deployment for compliance with the existing laws of war:

The fact that autonomous weapon systems will locate and attack persons and objects without human interaction raises unique issues. These challenges are not grounds for banning the systems entirely. On the contrary, the law of armed conflict's restrictions on the use of weapons (particularly the requirements that they be directed only against combatants and military objectives, that they not be employed indiscriminately, that their use not result in excessive harm to civilians or civilian objects, and that they not be used when other available weapons could achieve a similar military advantage while placing civilians and civilian objects at less risk) are sufficiently robust to safeguard humanitarian values. After all, as the ICRC [International Committee for the Red Cross] suggests, "new technologies do not change existing law, but rather must abide by it."<sup>36</sup>

#### **A Third Perspective**

While the starkly differing views presented by HRW, on the one hand, and the Reisner/Anderson/Waxman and Schmitt/Thurnher teams, on the other, might seem to occupy the full battlespace of legal and policy issues raised by LAWS, there is a third school of thought on this issue—one impliedly adopted by current DOD policy and endorsed by this paper. Under this third viewpoint, while LAWS do not—at least conceptually—represent an inherent violation of current international law, they do raise fundamentally new issues which are not fully addressed under current law. As such, under this assessment, the development and deployment of fully autonomous weapon systems should be held to an additional level of legal and policy constraints beyond compliance with the laws of war.

According to this view, autonomous weapon systems do not simply represent the latest development in the constant evolution of weaponry dating back to the first use of a rock or a club to defeat an adversary. Rather, they represent something conceptually and practically distinct from all the weapons that have come before them. As Christof Heyns, the United Nations' special rapporteur on extrajudicial, summary or arbitrary executions, has explained:

For societies with access to it, modern technology allows increasing distance to be put between weapons users and the lethal force they project. . . . Lethal Autonomous Robots (LARs) [Mr. Heyns' term for LAWS], if added to the arsenals of states, would add a new dimension to this distancing, in that targeting decisions could be taken by the robots themselves. . . . The robotics revolution has been described as the next major revolution

<sup>36.</sup> Schmitt and Thurnher, "Out of the Loop," 280.

in military affairs, on par with the introduction of gunpowder and nuclear bombs. But in an important respect LARs are different from these earlier revolutions: their deployment would entail not merely an upgrade of the kinds of weapons used, but also a change in the identity of those who use them. With the contemplation of LARs, the distinction between weapons and warriors risks becoming blurred, as the former would take autonomous decisions about their own use.<sup>37</sup>

Put more succinctly, for all their awesome power and lethality, nuclear weapons do not launch themselves.

As Mr. Heyns suggests, the distinctive character of autonomous weapon systems raises a host of novel risks and policy issues. The first is speed. As illustrated in Flash Boys, Michael Lewis's colorful history of the advent of highfrequency stock trading, research and development in modern computing has largely developed into a competition to shave off tiny increments of time, sometimes measured in picoseconds. In the race to beat their competitors to market, the traders described by Lewis, and the computer technicians and programmers supporting them, went to almost obscene lengths to ensure their orders would beat their competitors' to market in a race measured in increments incomprehensible to the human brain. In one instance, a particularly ambitious firm spent tens of millions of dollars buying land rights through rural Pennsylvania and Ohio in order to lay their own high-speed fiber optic cable. This dedicated cable ensured that the firm's buy and sell orders (determined by their equally speedy algorithms) could outrace their competitors' orders between the New York and Chicago stock exchanges in a race lasting less time than the blink of a human eye.<sup>38</sup>

The unprecedented speed with which autonomous systems of all types can determine and execute decisions presents significant issues when those decisions involve the use of deadly weapons. The speed of operations made possible by LAWS represents a paradigm shift in the conception of battle plan execution, which has throughout history unfolded no more quickly than the speed at which a human brain can assess information, weigh alternatives, and determine the best course of action. As Schmitt and Thurnher write, "Many nations, including China, are already developing advanced systems with autonomous features. Future combat may therefore occur at such a high tempo that human operators will simply be unable to keep up.... [Therefore] a force that does not employ fully autonomous weapon systems will inevitably operate outside its enemy's 'OODA [observe, orient, decide, act] loop,' thereby ceding initiative on the battlefield."<sup>39</sup>

<sup>37.</sup> Heyns, Report, 5-6.

<sup>38.</sup> Lewis, Flash Boys.

<sup>39.</sup> Schmitt and Thurnher, "Out of the Loop," 238.

The second distinctive feature of an autonomous weapon systems is, well, its autonomy. Never in the history of military affairs have weapons been designed with the potential not just to assist humans in *waging* wars, but to actually *commence* hostilities—even without a conscious human decision to do so.

This risk is compounded by the unpredictability of increasingly complex algorithms involving in some cases millions of lines of code. The more powerful the sensors connected to the platform, and the more factors the system's processor is asked to weigh before deciding on a course of action, the more difficult it is for human operators to predict how an autonomous system will react to any given sequence of real-world events or even to understand after the fact why the system reacted the way it did.<sup>40</sup> This risk factor is increased dramatically by the potential concurrent employment of autonomous systems by adversaries, each system reacting to the others' actions—or perceived actions—at a speed beyond the ability of the human brain to react to, or even comprehend, in real time.

The fourth and final novel risks posed by autonomous systems are hacking by adversaries and simple coding errors, either of which could lead to unanticipated and even deadly actions by improperly designed or secured LAWS. The US federal government's abysmal track record of acquiring and deploying IT systems, including the disastrous roll-out of the Affordable Care Act (Obamacare) marketplaces, the failure of data systems at agencies ranging from the Federal Bureau of Investigation to the Veterans Administration, and the highly publicized hacking of General Services Administration databases by Chinese actors, does not exactly inspire confidence in the government's ability to compile millions of lines of error-free code with impenetrable defenses against adversarial hacking.

The real-world dangers posed by the combination of the risk factors set forth above (speed, autonomous action, complexity, unknown errors, and hacking) are illustrated by Wall Street's famous "flash crash" of 6 May 2010. On that date, the Dow Jones Industrial Average suddenly lost nearly 10 percent of its total value in just minutes. As detailed by Mr. Scharre in his most recent paper on autonomous weapons, "A U.S. Securities and Exchange Commission (SEC) report following the incident determined that the crash was initiated by an automated stock trade (a 'sell algorithm') executing a large sale unusually quickly. This caused a sale that normally would have occurred over several hours to be executed within 20 minutes. This sell algorithm then interacted with high-frequency trading algorithms to cause a rapid price drop."<sup>41</sup>

<sup>40.</sup> See Office of the Chief Scientist, Autonomous Horizons, vol. 1, 5.

<sup>41.</sup> Scharre, "Autonomous Weapons," 35.

Perhaps most alarming of all, other organizations which have investigated the flash crash, including the Department of Justice and the Commodity Futures Trading Commission, have disputed the SEC's finding, instead attributing the flash crash to a London-based trader hacking into the autonomous trading algorithms of other firms. Scharre's conclusion, and the lack of certainty into the causes of the crash even seven years later, lends credence to the fears of many in the international community that autonomous weapon systems deployed into the wild without proper controls could act in unexpected and disastrous ways. "What appears clear across multiple analyses of the May 2010 incident is that automated stock trades and high-frequency trading algorithms at the very least played a role in exacerbating the crash. This may have been due, in part, to unanticipated interactions between adversarial trading algorithms. It is also possible that behavioral hacking of the algorithms was a factor."<sup>42</sup>

Given these risks, it is this paper's contention that simply applying the existing rule of law framework to these fundamentally novel systems is not sufficient to protect against the very real risks identified by HRW and others. Rather, additional standards are required to ensure that not only are the laws of armed conflict adhered to, but also that the principles behind those laws, including the protection of civilians and the prevention of unnecessary suffering by combatants and noncombatants alike, are respected in practice as well as in theory.

However, to say that there should be additional standards is a much simpler task than detailing what those standards should be. To identify proper standards for the development and use of autonomous weapon systems, it is necessary to not only identify the risks of such systems but also the reasons why advanced nations' militaries might wish to deploy LAWS outside the direct control of a human operator. In other words, why would a nation ever deploy a human out of the loop or fully autonomous system? As a review of the relevant literature indicates, there are three scenarios in which fully autonomous systems present distinct advantages over even semi-autonomous or human in the loop systems.<sup>43</sup> The first is when speed is required to execute or defend against an attack. In addition to the rapidly increasing speed of computer processors discussed above, the real-world velocity of propelled munitions and weapon systems is on the verge of a revolutionary increase with the anticipated deployment of hypersonic jets and munitions untethered from the limits of the human body to absorb g-forces. It is precisely this consideration that has led to the deployment of fully autonomous antiprojectile

<sup>42.</sup> Scharre, "Autonomous Weapons," 36.

<sup>43.</sup> See Office of the Chief Scientist, Autonomous Horizons, vol. 1, 1-2.

interceptor systems such as Israel's Iron Dome, the US Army's Patriot batteries, and the US Navy's Phalanx antimissile system for *Aegis*-class ships.<sup>44</sup> These systems are all designed to identify and intercept incoming missiles, mortars, and artillery rounds when speed is of the essence.

The second reason is related to the first, and that is to defend against a mass attack, whether from a large incoming missile volley or a future swarm attack by dozens, hundreds, or even thousands of small networked projectiles programmed to select and destroy targets in the most efficient manner possible.<sup>45</sup> Needless to say, choosing specific targets in fractions of a second from among potentially thousands of threats is well beyond the capacity of even the most highly trained human brain.

The final reason modern militaries may seek to deploy fully autonomous weapon systems is for operations in areas where communications are severely degraded or unavailable, whether because of system failure, jamming by an adversary, or to avoid detection of covert missions. Missions of this sort may involve strikes in antiaccess/area denial (A2/AD) areas such as the territory around the South China Sea or seek-and-destroy missions deep in enemy territory seeking predefined high value targets.

#### Meaningful Human Control

Within the international community, there is increasing support for the development of a global standard requiring meaningful human control of all weapon systems, including those of the autonomous variety.<sup>46</sup> At first blush, such a standard appears capable of ameliorating the unique risks inherent to LAWS without undermining their potential advantage in certain types of military conflict. Surely, one would hope, a weapon system under meaningful human control could not spontaneously set off the kinetic version of a flash crash. And yet, it would seem possible that a system under meaningful human control could still provide enhanced capabilities in the scenarios detailed above, for example by having a human on the loop of a defensive interceptor system or through human-machine teaming in a communications-denied or -degraded area.

<sup>44.</sup> For this discussion, I am adopting the HRW viewpoint that systems which have a human operator on the loop—but operate at speeds which make human intervention impractical, and therefore unlikely—are functionally indistinguishable from human out of the loop systems—and should thus be considered fully autonomous.

<sup>45.</sup> See Scharre, "Robotics on the Battlefield."

<sup>46.</sup> See Scharre and Horowitz, "Meaningful Human Control"; and Anderson, Reisner, and Waxman, "Adapting."

Indeed, without expressly endorsing the term, the DOD seems to be currently following a policy requiring meaningful human control in all but name. Department of Defense Directive (DODD) 3000.09, referenced above, states "it is DoD Policy that autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."<sup>47</sup> Additionally, the Directive defines fully autonomous weapons as those which "once activated, can select and engage targets without further intervention by a human operator" and forbids their development unless approved by two undersecretaries of defense and the chairman of the Joint Chiefs of Staff before development and again before fielding. To ensure that human control is meaningful in practice as well as in theory, the directive requires that the interfaces of any autonomous or semi-autonomous systems: "(a) Be readily understandable to trained operators, (b) Provide traceable feedback on system status, and (c) Provide clear procedures for trained operators to activate and deactivate system functions."<sup>48</sup>

The US Air Force's recent publication of *Autonomous Horizons*, vol. 1 sets a vision which is entirely consistent with both DODD 3000.09 and the notional international standard of meaningful human control. Indeed, even volume 1's subtitle, *Human-Autonomy Teaming*, is suggestive of the meaningful human control standard. The guide goes on to set forth a vision of autonomous systems supporting and complementing, rather than replacing, the ultimate decision making of the human operator:

Autonomous systems will be designed to serve as part of a collaborative team with airmen. Flexible autonomy will allow the control of tasks, functions, sub-systems, and even entire vehicles to pass back and forth over time between the airman and the autonomous system, as needed to succeed under changing circumstances.... In certain limited cases the system may allow the autonomy to take over automatically from the airman, when timelines are very short for example or when loss of lives are [*sic*] imminent. However, human decision making for the exercise of force with weapon systems is a fundamental requirement, in keeping with the Department of Defense directives.<sup>49</sup>

While the meaningful human control standard thus has wide, if not unanimous, support from within the international community and (in principle if not in name) from the DOD and the USAF, it ultimately falls prey to the same definitional debates that characterize so much of the discussion on LAWS. What is meaningful human control? Does it require a human to physically operate the system, or does real-time supervision suffice? Must the human operator be in the loop approving targets, or would the ability to intervene to

<sup>47.</sup> DODD 3000.09, 2.

<sup>48.</sup> DODD 3000.09, 2.

<sup>49.</sup> USAF Office of the Chief Scientist, Autonomous Horizons, vol. 1, v.

countermand a target selection (on the loop) meet the standard? For that matter, could a human out of the loop or fully autonomous system meet the standard of meaningful human control if it operates under sufficiently narrow parameters set in advance by a human operator—for example to search for and engage one specific target that had been approved prior to deployment by the appropriate human authority? Needless to say, there is nothing approaching consensus on any of these questions, making the meaningful human control standard something less than meaningful in practice.

As Mr. Scharre has observed following recent discussions at the United Nations (UN),

There is no clear definition or agreement at this point, although, as the UN Institute for Disarmament Research points out, "the idea of Meaningful Human Control is intuitively appealing even if the concept is not precisely defined." Without a clear definition, however, 'meaningful human control' risks being only a pleasant-sounding catchphrase. At best, it merely shifts the debate to, "what is meaningful?" It also risks appearing to resolve the issues raised by increased autonomy in weapons, when in reality it becomes an empty platitude, and one that would be devoid of a common meaning. At worst, a failure to define the term clearly could, if embedded in international discussions, lead to flawed policy choices.<sup>50</sup>

Any attempts to more concretely define meaningful human control must be substantive while providing sufficient flexibility to account for the development of increasingly sophisticated future systems, current and historical standards for the use of weapons in combat, and the wide variety of contexts in which LAWS might be used. A recent attempt at a definition by the International Committee on Robotic Arms Control (ICRAC) failed on all fronts. As described by Scharre, the 2014 definition of meaningful human control proposed by the ICRAC includes a provision that, for meaningful human control to be exercised, a commander must have "full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack." Yet, humans have been employing weapons where they lack perfect, real-time situational awareness of the target area since at least the invention of the catapult.

The ICRAC statement also argues, "There must be a means for the rapid suspension or abortion of the attack." However, the essence of a projectile weapon, since the first time a human hurled a rock in anger, is the inability to suspend and abort the attack after launch. It is only with modern, advanced weapons that commanders have had the ability to retarget or abort a projectile in flight. These proposed requirements articulate an idealized version of

<sup>50.</sup> Scharre and Horowitz, "Meaningful Human Control," 5.

human control divorced from the reality of warfare and the weapons that have long been considered acceptable in conducting it.

Not only would they arguably outlaw whole classes of weapons dating back to the sling and stone, but they also fail to capture what is new about autonomy. If meaningful human control is defined in such a way that it has never existed in war, or only very rarely, then such a definition sheds little light on the new challenge posed by increased autonomy.<sup>51</sup>

Any useful definition of meaningful human control must also avoid resorting to black and white binary categorizations of actions and circumstances that often fall into a discomfiting gray area. For example, DODD's definition of a fully autonomous system as one which "once activated, can select and engage targets without further intervention by a human operator" raises more questions than it answers. In a future world of swarm versus swarm battles, or when facing an incoming artillery barrage, what does it mean for a human to "select" a target? Does the human operator have to specifically instruct each interceptor which projectile it should aim for? Is it enough to point the system to a general area from which the threats are emanating and allow the system to fashion the best defense? For that matter, *when* must the human select the target? If a human operator programs an autonomous system to seek out and destroy a specific target, including a human being, and the system then executes that mission hours, days, or weeks later, has that target still been selected by a human being?

Similarly, while comforting, it is difficult to distinguish between offensive and defensive uses of autonomous systems. While current autonomous projectile interceptors such as the Iron Dome, Phalanx, and Patriot systems have raised little concern within the international community, including HRW, because of their fundamentally defensive character, defining offensive versus defensive operations has been a famously contentious and fraught exercise throughout the modern history of warfare. Imagine a system which, after detecting a hostile missile or artillery launch, rather than firing at the incoming projectiles instead fires at the launchers themselves, even if they are located in enemy sovereign territory. What if it fires after detecting an "imminent" launch, but before the actual launch? What if the system is deployed in an area as large as the South China Sea to autonomously attack "enemy" weapons platforms which "invade" contested territory over which China claims sovereignty? As these examples illustrate, any refinement of the meaningful human control standard must avoid categorical distinctions which will almost certainly fail to do anything other than provide the vocabulary for future disputes.

<sup>51.</sup> Scharre and Horowitz, "Meaningful Human Control," 9.

Instead, this paper proposes the adoption of a three-factor test for determining whether any given system is operating under "meaningful human control." In using this standard, lawyers and policy makers performing the mandatory legal review of new weapon systems both at the development and the deployment phase would assess: (1) the length of the potential lapse of time between the loss of direct human control and the weapon's latest possible engagement with a target, (2) the potential geographic area in which the weapon may engage a target, and (3) the degree of discernment the system is likely to use prior to engaging the target.<sup>52</sup> These three factors would be evaluated collectively, such that a system which is evaluated poorly under one or two of the factors must score highly in the remaining factor or factors if it is to be determined to be within meaningful human control.

The first factor, duration, would assess how much time might lapse between the loss of direct human control of the weapon (defined as the ability of a human operator to abort or divert any planned or authorized attack) and the last moment the system would be authorized to, or capable of, striking a target. Under this standard, then, a fire-and-forget air-to-air missile which can only chase a target for several minutes would be considered more "under control" than the Defense Advanced Research Projects Agency's (DARPA) planned Anti-Submarine Warfare Continuous Trail Unmanned Vessel (AC-TUV), which is reported to have the ability to hunt enemy submarines for up to 90 days without human contact.<sup>53</sup>

The second factor, geographic range, would look to not only the size of the box in which the system is authorized to operate (or is capable of traveling), but also the nature of the territory. A system which is authorized to operate over a large area consisting exclusively of open seas, or uninhabited desert, might pose less risk, and thus score higher, than a system designed to operate within several city blocks of a densely populated urban environment.

Finally, the third factor would look at the system's capacity for, and requirement of, discernment before engaging a target. This analysis can be thought of as a range between, on the low end, a "dumb" weapon such as an antipersonnel land mine that will blow up anything that triggers it, to, on the high end, a search-and-destroy unmanned aerial vehicle programmed to look for one specific high value target and to only strike if it can (1) positively identify the target using its sensors and (2) determine that any collateral damage will be

<sup>52.</sup> The term "discernment" used in this paper is intended to be broader than the term "distinction" used in international law since, ideally, autonomous systems would distinguish among valid military targets, and the circumstances surrounding a strike on such targets, and not just between military targets and banned civilian targets.

<sup>53.</sup> Schmitt and Thurnher, "Out of the Loop," 240.

within acceptable standards (as determined by the appropriate human authority prelaunch). Falling somewhere in the middle would be guided ordnance programmed to strike set coordinates and interceptor missiles designed to engage incoming targets fitting a certain signature.

Applying this analysis to a range of weapons, real or envisioned, which are today thought of as either autonomous or semi-autonomous, illustrates how such a test might work in practice. An antipersonnel land mine, often considered the most basic form of autonomous weapon in that it does "select" its target without human intervention, would fare poorly on the first and third factors, with the second question dependent on the circumstances of use. Since a land mine does not "turn off" on its own or check in with a human operator, the lapse of time between its deployment and potential engagement with a target is virtually infinite. In fact, the more than 23,000 deaths in Afghanistan since 1979 caused by the more than 600,000 land mines placed in that country by the Soviets during the Afghan–Soviet war speak to the enduring dangers represented by land mines years or decades after their placement.<sup>54</sup>

On the second factor, while the geographic range of a particular land mine may be no more than a few feet, land mines are usually placed in bunches, over large swathes of territory. The placement of land mines in areas likely to be frequented by civilians exacerbates this issue, while placements limited to areas off-limits or inaccessible to civilians, such as the demilitarized zone between the two Koreas, significantly mitigates the threat to the civilian population and would increase the mines' "score" on the three-factor test.

No matter where they are placed, land mines would score very low on the third factor, discernment. Lacking the ability to distinguish between an enemy soldier on a military mission, a child playing soccer, or a grazing goat, antipersonnel mines are the ultimate dumb weapons. Antivehicle mines, which, if properly employed, can only be set off by heavy vehicles such as tanks, would score somewhat higher, though they still lack the ability to distinguish between friend and foe or, for that matter, between a tank and a large truck. Thus, unless carefully employed only in areas devoid of civilians, and then removed as soon as the conflict is over or civilians begin entering the area, antipersonnel land mines are likely to fail the test of meaningful human control and would thus be banned under the proposed three-factor test.

Scoring much higher on the three-factor test would be a guided weapon such as Raytheon's advanced medium-range air-to-air missile (AMRAAM). Though as a fire-and-forget weapon, the pilot (or weapon systems officer) loses control of the missile once deployed, the AMRAAM has a maximum

<sup>54.</sup> Halo Trust, "Where We Work."

mission duration of only several minutes from launch to detonation. On the second factor, the missile is used only in a limited geographic range (within approximately 100 km of launch depending on the model) and is used only in combat environments thousands of feet in the air. Finally, with guidance systems dictated first by the launching plane's radar and other sensors, and then, in its terminal phase, by the munition's internal radar, the AMRAAM is designed to track and strike only the designated target and poses little threat of civilian death or destruction.<sup>55</sup>

A future system such as DARPA's ACTUV unmanned submarine, described above, would likely score somewhere between the land mine and the AMRAAM, depending on its exact specifications and terms of use. With potential deployment of up to 90 days in the deep sea without human interface, the lapse of time between loss of human control and potential target engagement is certainly much longer than seen with the AMRAAM, though much less egregious than the land mines in Afghanistan. Given their long deployment cycles, the ACTUV could also potentially cover huge stretches of open ocean, including those frequented by civilian vessels (though this is much less likely if their deployment is limited to the deep sea). The capacity for discernment is unknown at this point, though it would presumably be intended to operate under very specific and limited target parameters, including the capacity to credibly identify enemy submarines through physical, electronic, or other signatures detectable through the ACTUV's sensors. Concerns about the ACTUV as deployed could be significantly ameliorated by requiring it to surface and check in with a human operator on a regular basis, limiting its patrol area, or narrowing its targeting parameters.

The three-factor test detailed above is certainly not without its drawbacks. As some in the skeptic community would doubtless point out, the proposed test would result in a significant gray area which nations would likely exploit to their advantage. How to assess each proposed weapon under each factor, and how to weigh the factors against each other, could lead to endless debate and make prosecutions or sanctions for claimed violations difficult to sustain. While containing some merit, this complaint is not unique to LAWS. The traditional laws of war rely on equally nebulous principles, including the weighing of military advantage against likely collateral damage under the principle of proportionality. Moreover, virtually every legal system in the world relies on similar multifactor balancing tests in a variety of complex or emerging fields, meaning lawyers and judges have become adept at precisely this type of analysis. Ultimately, any legal standard that attempts to balance the

<sup>55.</sup> Forecast International, "About the AIM 120 AMRAAM."

provision of substantive guidance with the necessity to provide flexibility in the face of evolving technologies, and the limitless range of possible contingencies where LAWS may be used, will fail to satisfy skeptics seeking a bright line rule. The international community should not allow this pursuit of purity to prevent international law from developing to account for the new and unique challenges brought on by the advent of autonomous weapon systems.

#### Conclusion

The debate over the rules and standards that do, or should, apply to lethal autonomous weapon systems is still in its nascent stage, as are the systems themselves. While this debate will doubtless continue and take as many unexpected turns as the technology itself, it is vitally important that the major players in the development of such systems engage seriously in the discussion, with an end to reaching agreement on at least broad standards before the technology, or the outset of unexpected conflict, eclipses any such efforts. While the three-factor test proposed is unlikely to settle the debate, it is an attempt to accommodate the realities of modern warfare and the uncertainties surrounding emerging technologies with the need to provide additional substance to the current active or proposed standards. While the goal of "civilizing" the inherently brutal and inhuman nature of warfare has always been tinged with an idealism decried by its critics, it is precisely these international efforts that have allowed society to benefit from the incredible potential of modern systems to minimize civilian casualties and unnecessary suffering while constraining the potentially awesome power inherent in these same systems. If we are to continue to abide by the principles and ideals embodied in the Geneva and Hague conventions and similar international agreements, the efforts at the national and international levels to modernize the rules behind these principles must continue.

### Bibliography

- Anderson, Kenneth, Daniel Reisner, and Matthew Waxman. "Adapting the Law of Armed Conflict to Autonomous Weapon Systems." *International Law Studies*, no. 90 (2014): 386–411. https://papers.ssrn.com.
- Department of Defense (DOD). DOD Directive 3000.09, Autonomy in Weapon Systems. 12 November 2012.
- Forecast International. "About the AIM 120 AMRAAM." AEROWEB. 5 June 2015. http://www.fi-aeroweb.com/Defense/AMRAAM.html.
- Halo Trust. "Where We Work." www.halotrust.org.
- Heyns, Christof. "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions." United Nations General Assembly: Human Rights Council, April 2013. hchr.org.
- Human Rights Watch and The International Human Rights Clinic. Losing Humanity: The Case Against Killer Robots. 19 November 2012. www.hrw.org.
- Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I). International Committee of the Red Cross. 8 June 1977. https:// ihl-databases.icrc.org.
- Lewis, Michael. *Flash Boys: A Wall Street Revolt.* New York: W. W. Norton and Company Inc., 2014.
- Scharre, Paul. "Autonomous Weapons and Operational Risk." Center for a New American Security. 29 February 2016. www.cnas.org.
  - ——. "Robotics on the Battlefield Part II: The Coming Swarm." Center for a New American Security. 15 October 2014. www.cnas.org.
- Scharre, Paul, and Michael C. Horowitz. "An Introduction to Autonomy in Weapon Systems." Center for a New American Security. 13 February 2015. www.cnas.org.
  - ——. "Meaningful Human Control in Weapon Systems: A Primer." Center for a New American Security. 16 March 2015. www.cnas.org.
- Schmitt, Michael N., and Jeffrey S. Thurnher. "Out of the Loop:' Autonomous Weapon Systems and the Law of Armed Conflict." *Harvard National Security Law Journal*, vol. 4 (2013). https://harvardnsj.org.
- USAF Office of the Chief Scientist. *Autonomous Horizons: System Autonomy in the Air Force—A Path to the Future*, vol. 1: *Human-Autonomy Teaming* (June 2015). www.af.mil.