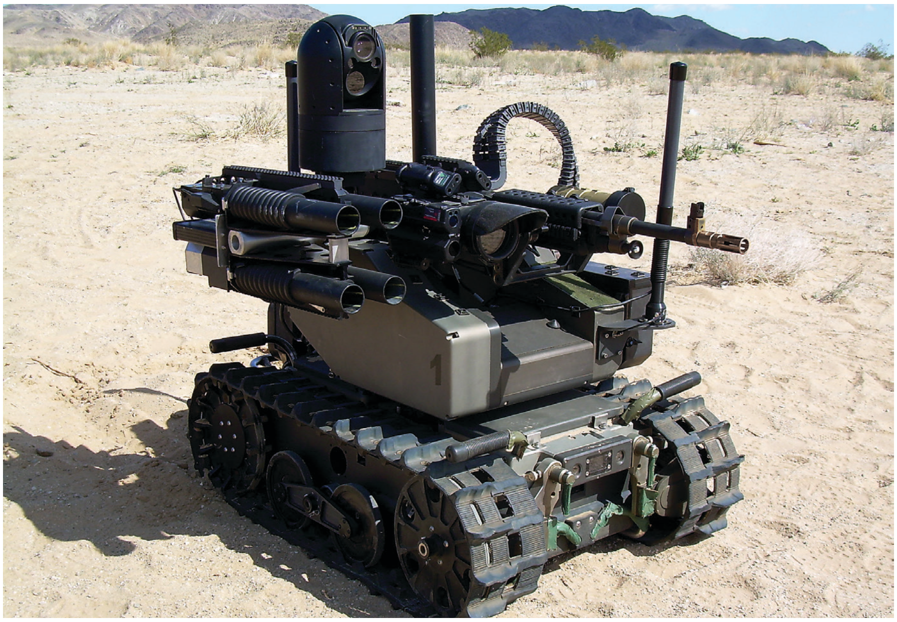
DOI:10.1145/2998579

Wendell Wallach

Viewpoint
Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles

A 10-point plan toward fashioning a proposal to ban some—if not all—lethal autonomous weapons.



The Modular Advanced Armed Robotic System is an unmanned ground vehicle for reconnaissance, surveillance, and target acquisition missions.

IMAGE COURTESY OF QINETIQ NORTH AMERICA

FROM APRIL 11–15, 2016, at the United Nations Office at Geneva, the Convention on Certain Conventional Weapons (CCW) conducted a third year of informal meetings to hear expert testimony regarding a preemptive ban on lethal autonomous weapons systems (LAWS). A total of 94 states attended the meeting, and at the end of the week they agreed by consensus to recommend the formation of an open-ended Group of Government Experts (GGE). A GGE is the next step in forging a concrete proposal upon which the member states could vote. By the end of 2016 a preemptive ban has been called for by 19 states. Furthermore, *meaningful human control*, a phrase first proposed by advocates for a ban, has been adopted by nearly all the states, although the phrase's meaning is contested. Thus a ban on LAWS would appear to have gained momentum. Even the large military powers, notably the U.S., have publicly stated that they will support a ban if that is the will of the member states. Behind the scenes, however, the principal powers express their serious disinclination to embrace a ban. Many of the smaller states will follow their lead. The hurdles in the way of a successful campaign to ban LAWS remain daunting, but are not insurmountable.

The debate to date has been characterized by a succession of arguments and counterarguments by proponents and opponents of a ban. This back and forth should not be interpreted as either a stalemate or a simple calculation as to whether the harms of LAWS can be offset by their benefits. For all states that are signatories to the laws of armed conflict,[a] any violation of the principles of international humanitarian law (IHL)[b] must trump utilitarian calculations. Therefore, those who believe the benefits of LAWS justify their use and therefore oppose a ban, are intent that LAWS do not become a special case within IHL. Demonstrating that LAWS pose unique challenges

[a] LOAC, also known as International Humanitarian Law (IHL), is codified in the Geneva Conventions and additional Protocols. The laws seek to limit the effects of armed conflict, particularly the protection of non-combatants.

[b] Four principles of IHL provide protection for civilians: distinction, necessity and proportionality, humane treatment, and nondiscrimination.

for IHL has been a core strategy for supporters of a ban.

Those among the more than 3,100 AI/Robotics researchers who signed the *Autonomous Weapons: An Open Letter From AI & Robotics Researchers*[c] are reflective of a broad consensus among citizens and even active military personnel who favor a preemptive ban.[4] This consensus is partially attributable to speculative, futuristic, and fictional scenarios. But perhaps even science fiction represents a deep intuition that unleashing LAWS is not a road humanity should tread.

Researchers who have waded into the debate over banning LAWS have come to appreciate the manner in which geopolitics, security concerns, the arcana of arms control, and linguistic obfuscations can turn a relatively straightforward proposal into an extremely complicated proposition. A ban on LAWS does not fit easily, or perhaps at all, into traditional models for arms control. If a ban, or even a moratorium, on the development of LAWS is to progress, it must be approached creatively.

I favor and have been a long-time supporter of a ban. While a review of the extensive debate as to whether LAWS should be banned is well beyond the scope of this paper, I wish to share a few creative proposals that could move the campaign to ban LAWS forward. Many of these proposals were expressed during my testimony at the CCW meeting in April and during a side luncheon event.[d] Before introducing those proposals, let me first point out some of the obstacles to fashioning an arms control agreement for LAWS.

Why Banning LAWS Is Problematic

▸ Unlike most other weapons that have been banned, some uses of LAWS are perceived as morally acceptable, if not morally obligatory. The simple fact that LAWS can be substituted for and thus save the lives of one's own soldiers is the most obvious moral good. Unfortunately, this same moral good lowers the barriers to initiating new wars. Some nations will be emboldened to start wars if they believe they can achieve political objectives without the loss of their troops.

Some nations will be emboldened to start wars if they believe they can achieve political objectives without the loss of their troops.

▸ It is unclear whether armed military robots should be viewed as weapon systems or weapon platforms, a distinction that has been central to many traditional arms control treaties. Range, payload, and other features are commonly used in arms control agreements to restrict the capabilities of a weapon system. A weapon platform can be regulated by restricting where it can be located. For example, agreements to restrict nuclear weapons will specify number of warheads and the range of the missiles upon which they are mounted, and even where the missiles can be stationed. With LAWS, what is actually being banned?

• Arms control agreements often focus on working out modes of verification and inspection regimes to determine whether adversaries are honoring the ban. The difference between a lethal and non-lethal robotic system may be little more than a few lines of code or a switch, which would be difficult to detect and could be removed before or added after an inspection. Proposed verification regimes for LAWS[6] would be extremely difficult and costly to enforce. Military strategists do not want to restrict their options, when that of bad actors is unrestricted.

• LAWS differ in kind from the various weapon systems that have to date been banned without requiring an inspection regime. Consider, for example, the relatively recent bans on blinding lasers or anti-personnel weapons, which are often offered as a model for arms control for LAWS. These bans rely on representatives of *civil society*, non-governmental organizations such as the International Committee of the Red Cross, to monitor and stigmatize violations. So also will a ban on LAWS. However, blinding lasers and anti-personnel weapons were relatively easy to define. After the fact, the use of such weapons can be proven in a straightforward manner. Lethal autonomy, on the other hand, is not a weapon system. It is a feature set that can be added to many, if not all, weapon systems. Furthermore, the uses of autonomous killing features are likely to be masked.

• LAWS will be relatively easy to assemble using technologies developed for civilian applications. Thus their proliferation and availability to non-state actors cannot be effectively stopped.

In forging arms-control agreements definitional distinctions have always been important. Contentions that definitional consensus cannot be reached for *autonomy* or *meaningful human control*, that LAWS depend upon advanced AI, and that such systems are merely a distant speculative possibility repeatedly arose during the April discussion at the U.N. in Geneva, and generally served to obfuscate, not clarify, the debate. A circular and particularly unhelpful debate has ensued over the meaning of *autonomy*, with proponents and opponents of a ban struggling to establish a definition that serves their cause. For example, the U.K. delegation insists that *autonomy* implies near humanlike capabilities[e] and anything short of this is merely an *automated* weapon. The Campaign to Stop Killer Robots favors a definition where *autonomy* is the ability to perform a task without immediate intervention from a human. Similarly, definitions for *meaningful human control* range from a military leader specifying a kill order in advance of deploying a weapon system to having the real-time engagement of a human *in the loop* of selecting and killing a human target.

[c] Available at http://bit.ly/1V9bls5

[d] The full April 12, 2016, testimony entitled, Predictability and Lethal Autonomous Weapons Systems (LAWS), is available at http://bit.ly/2mjmuwH. An extended article accompanied this testimony. That article was circulated to all the CCW member states by the chair of the meeting, Ambassador Michael Biontino of Germany. It was also published in Robin Geiss, Ed., 2017, "Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security." Federal Foreign Office, p. 295–312. The luncheon event on April 11, 2016, was sponsored by the United Nations Institute for Disarmament Research (UNIDIR).

[e] While the U.K. representatives did not use this language, it does succinctly capture the delegation's statements that all computerized systems are merely automated until they display advanced capabilities.

The leading military powers contend that they will maintain effective control over the LAWS they deploy.[f] But even if we accept their sincerity, this totally misses the point. They have no means of ensuring that other states and non-state actors will follow suit.

More is at stake in these definitional debates than whether to preemptively ban LAWS. Consider a Boston Dynamic's Big Dog loaded with explosives, and directed through the use of a GPS to a specific location, where it is programmed to explode. Unfortunately, during the time it takes to travel to that location, the site is transformed from a military outpost to a makeshift hospital for injured civilians. A strong definition for *meaningful human control* would require the location be given a last-minute inspection before the explosives could detonate. Big Dog, in this example, is a dumb LAW, which we should perhaps fear as much as speculative future systems with advanced intelligence. Dumb LAWS, however, do open up comparisons to widely deployed existing weapon systems, such as cruise missiles, whose impact on an intended target military leaders have little or no ability to alter once the missile has been launched. In other words, banning dumb LAWS quickly converges with other arms control campaigns, such as those directed at limiting cruise missiles and ballistic missiles.[5] States will demand a definition for LAWS that distinguishes them from existing weapon systems.

Delegates at the CCW are cognizant that in the past (1990s) they failed at banning the dumbest, most indiscriminate, and autonomous weapons of all, anti-personnel mines. Nevertheless, anti-personnel weapons (land mines) were eventually banned during an independent process that led up to the Mine Ban or Ottawa Treaty; 162 countries have committed to fully comply with that treaty.[g]

The leading military powers contend they will maintain effective control over the LAWS they deploy. But even if we accept their sincerity, this totally misses the point.

A second failure to pass restrictions on the use of a weapon systems, whose ban has garnered popular support, might damage the whole CCW approach to arms control. This knowledge offers the supporters of a ban a degree of leverage presuming: the ban truly has broad and effective public support; LAWS can be distinguished from existing weaponry that is widely deployed; and creative means can be forged to develop the framework for an agreement.

A 10-Point Plan

Many of the barriers to fitting a ban on LAWS into traditional approaches to arms control can be overcome by adopting the following approach.

1. Rather than focus on establishing a bright line or clear definition for *lethal autonomy*, first establish a high order moral principle that can garner broad support. My candidate for that principle is: *Machines, even semi-intelligent machines, should not be making life and death decisions. Only moral agents should make life and death decisions about humans.* Arguably, something like this principle is already implicit, but not explicit, in existing international humanitarian law, also known as the laws of armed conflict (LOAC).[3] A higher order moral principle makes explicitly clear what is off limits, while leaving open the discussion of marginal cases where a weapon system may or may not be considered to be *making life and death decisions*.

2. Insist that *meaningful human control* and *making a life and death decision* requires the real-time authorization from designated military personnel for a LAW to kill a combatant or destroy a target that might harbor combatants and non-combatants alike. In other words, it is not sufficient for military personnel to merely delegate a kill order in advance to an autonomous weapon or merely be "on-the-loop"[h] of systems that can act without a real time go-ahead.

3. Petition leaders of states to declare that LAWS violate existing IHL. In the U.S. this would entail a Presidential Order to that effect.[i,14]

4. Review marginal or ambiguous cases to set guidelines for when a weapon system is truly autonomous and when its actions are clearly the extension of a military commander's will and intention. Recognize that any definition of autonomy will leave some cases ambiguous.

5. Underscore that some present and future weapon system will occasionally act unpredictably and most LAWS will be difficult if not impossible to test adequately.

6. Present compelling cases for banning at least some, if not all, LAWS. In other words, highlight situations in which nearly all parties will support a ban. For example, no nation should want LAWS that can launch nuclear warheads.

7. Accommodate the fact that there will be necessary exceptions to any ban. For example, defensive autonomous weapons that target unmanned incoming missiles are already widely deployed.[j] These include the U.S. Aegis Ballistic Missile Defense System and Israel's Iron Dome.

8. Recognize that future technological advances may justify additional

[f] See, for example, the U.S. Department of Defense Directive 2000.09 entitled, "Autonomy in Weapon Systems." The Directive is dated November 21, 2012 and signed by Deputy Secretary of Defense, Ashton B. Carter, who was appointed Secretary of Defense by President Obama on December 5, 2014; http://bit.ly/1myJikF

[g] The U.S., Russia, and China are not signatories to the Ottawa Treaty, although the U.S. has pledged to largely abide by its terms.

[h] "On the loop" is a term that first appeared in the "United States Air Force Unmanned Aircraft Systems Flight Plan 2009–2047." The plan states: Increasingly humans will no longer be "in the loop" but rather "on the loop"—monitoring the execution of certain decisions. Simultaneously, advances in AI will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input.

[i] Wallach, W. (2012, unpublished but widely circulated proposal). Establishing limits on autonomous weapons capable of initiating lethal force.

[j] In practice a weapon designed for defensive purposes might be used offensively. So the distinction between the two should emphasize the use of defensive weaponry to target unmanned incoming missiles.

exceptions to a ban. Probably the use of LAWS to protect refugee non-combatants would be embraced as an exception. Whether the use of LAWS in a combat zone where there are no non-combatants should be treated as an exception to a ban would need to be debated. Offensive autonomous weapon systems that do not target humans, but only target, for example, unmanned submarines, might be deemed an exception.

9. Utilize the unacceptable LAWS to campaign for a broad ban, and a mechanism for adding future exceptions.

10. Demand that the onus of ensuring that LAWS will be controllable, and that those who deploy the LAWS will be held accountable, lies with those parties who petition for, and deploy, an exception to the ban.

Unpredictable Behavior: Why Some LAWS Must Be Banned

A ban will not succeed unless there is a compelling argument for restricting at least some, if not all, LAWS. In addition to the ethical arguments for and against LAWS, concern has been expressed that autonomous weapons will occasionally behave unpredictably and therefore might violate IHL, even when this is not the intention of those who deploy the system. The ethical arguments against LAWS have already received serious attention over the past years and in the ACM. During my testimony at the CCW in April 2016, I fleshed out why the prospect of unanticipated behavior should be taken seriously by member states. The points I made are fairly well understood within the community of AI and robotics' engineers, and go beyond weaponry to our ability to predict, test, verify, validate, and ensure the behavior and reliability of software and indeed any complex system. In addition, debugging and ensuring that software is secure can be a costly and a never-ending challenge.

Factors that influence a system's predictability. Predictability for weaponry means that within the task limits for which the system is designed, the anticipated behavior will be realized, yielding the intended result. However, nothing less than a law of physics is absolutely predictable. There are only degrees of predictability, which in theory can be represented as a probability. Many factors influence the predictability of a system's behavior, and whether operators can properly anticipate the system's behavior.

- An unanticipated event, force, or resistance can alter the behavior of even highly predictable systems.
- Many if not most autonomous systems are best understood as complex adaptive systems. Within systems theory, complex adaptive systems act unpredictably on occasion, have tipping points that lead to fundamental reorganization, and can even display emergent properties that are difficult, if not impossible, to explain.
- Complex adaptive systems fail for a variety of reasons including incompetence or wrongdoing; design flaws and vulnerabilities; underestimating risks and failure to plan for low probability events; unforeseen high-impact events (Black Swans;[12] and what Charles Perrow characterized as uncontrollable and unavoidable "normal accidents" (discussed more fully here).
- Reasonable testing procedures will not be exhaustive and can fail to ascertain whether many complex adaptive systems will behave in an uncertain manner. Furthermore, the testing of complex systems is costly and only affordable by a few states, and they tend to be under pressure to cut military expenditures. To make matters worse, each software error fixed and each new feature added can alter a system's behavior in ways that can require additional rounds of extensive testing. No military can support the time and expense entailed in testing systems that are continually being upgraded.
- Learning systems can be even more problematic. Each new task or strategy learned can alter a system's behavior and performance. Furthermore, learning is not just a process of adding and altering information; it can alter the very algorithm that processes the information. Placing a system on the battlefield that can change its programming significantly raises the risk of uncertain behavior. Retesting dynamic systems that are constantly learning is impossible.
- For some complex adaptive systems various mathematical proofs or formal verification procedures have been used to ensure appropriate behaviors. Existing approaches to formal verification will not be adequate for

Calendar of Events

May 6–11
**CHI'17: CHI Conference on Human Factors in Computing Systems,
Denver, CO,
Sponsored: ACM/SIG,
Contact: Susan R. Fussell,
Email: sfussell@cornell.edu**

May 8–10
**HotOS '17: Workshop on Hot Topics in Operating Systems,
Whistler, BC, Canada,
Sponsored: ACM/SIG,
Contact: Rachit Agarwal,
Email: ragarwal@cs.cornell.edu**

May 10–12
**GLSVLSI '17: Great Lakes Symposium on VLSI 2017,
Banff, AB, Canada,
Contact: Laleh Behjat,
Email: laleh@ucalgary.ca**

May 15–17
**CF'17: Computing Frontiers Conference,
Siena, Italy,
Sponsored: ACM/SIG,
Contact: Roberto Giorgi,
Email: giorgi@unisi.it**

May 22–24
**SYSTOR 2017: International Systems and Storage Conference,
Haifa, Israel,
Sponsored: ACM/SIG,
Contact: Doron Chen,
Email: cdoron@il.ibm.com**

May 24–26
**SIGSIM-PADS '17: SIGSIM Principles of Advanced Discrete Simulation
Singapore, Singapore,
Sponsored: ACM/SIG,
Contact: Wentong Cai,
Email: aswtcai@ntu.edu.sg**

June

June 5–7
**Web3D '17: The 22nd International Conference on Web3D Technology,
Brisbane, QLD, Australia,
Sponsored: ACM/SIG,
Contact: Matt Adcock,
Email: matt.adcock@csiro.au**

June 5–8
**ICMR '17: International Conference on Multimedia Retrieval,
Bucharest, Romania,
Sponsored: ACM/SIG,
Contact: Niculae Sebe,
Email: sebe@disi.unitn.it**

systems with learning or planning capabilities functioning in complex socio-technical contexts. However, new formal verification procedures may be developed. The success of these will be an empirical question, but ultimately political leaders and military planners must judge whether such approaches are adequate for ensuring that LAWS will act within the constraints of IHL.

▸ While increasing autonomy, improving intelligence, and machine learning can boost the system's accuracy in performing certain tasks; they can also increase the unpredictability in how a system performs overall.

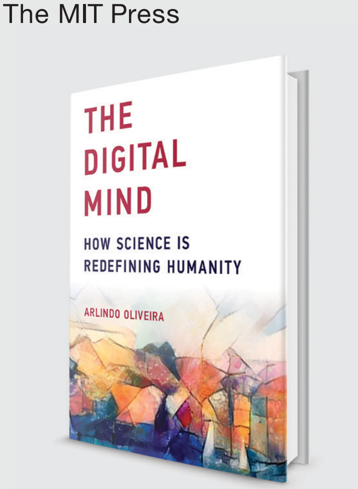
▸ Unpredictable behavior from a weapon system will not necessarily be lethal. But even a low-risk autonomous weapon will occasionally kill non-combatants, start a new conflict, or escalate hostilities.

Coordination, Normal Accidents, and Trust. Military planners often underestimate the risks and costs entailed in implementing weapon systems. Analyses often presume a high degree of reliability in the equipment deployed, and ease at integrating that equipment into a combat unit. Even autonomous weapons will function as components within a team that will include humans fulfilling a variety of roles, other mechanical or computational systems, and an adequate supply chain serving combat and non-combat needs.

Periodic failures or system accidents are inevitable for extremely complex systems. Charles Perrow labeled such failures "normal accidents."[8] The near meltdown of a nuclear reactor at Three Mile Island in Pennsylvania on March 28, 1979, is a classical example of a normal accident. Normal accidents will occur even when no one does anything wrong. Or they can occur in a joint cognitive system—where both operators and software are selecting courses of action—when it is impossible for the operators to know the appropriate action to take in response to an unanticipated event or action by a computational system. In the latter case, the operators do the wrong thing, because they misunderstand what the semi-intelligent system is trying to do. This was the case on December 6, 1999, when after a successful landing, confusion reigned, and a Global Hawk unmanned air vehicle veered off the runway and its nose collapsed in the adjacent desert, incurring $5.3 million in damages.[7]

In a joint cognitive system, when anything goes wrong, the humans are usually judged to be at fault. This is largely because of assumptions that the actions of the system are automated, while humans are presumed to be the adaptive players on the team. A commonly proposed solution to the failure of a joint cognitive system will be to build more autonomy into the computational system. This strategy, however, does not solve the problem. It becomes ever more challenging for a human operator to anticipate the actions of a smart system, as the system and the environments in which it operates become more complex. Expecting operators to understand how a sophisticated computer *thinks*, and to anticipate its actions so as to coordinate the activities of the team, increases the responsibility of the operators.

Difficulty anticipating the actions of other team members (human or computational) in turn undermines trust, an essential and often overlooked element of military preparedness. Heather Roff and David Danks



have analyzed the challenges entailed for ensuring that human combatants will *trust* LAWS. For autonomous weaponry that have either planning capabilities or learning capabilities, they conclude that ensuring trust will require significant time, training, and cost.[10] This certainly does not rule out a satisfactory integration of LAWS into combat units. But it does suggest resources and costs that are seldom factored into determinations that autonomous systems are cost effective. Furthermore, there should be concerns as to whether appropriate training for combatants working with LAWS will actually be provided.

Since Perrow first proposed his theory of *normal accidents*, it has been fleshed out into a robust framework for understanding the safety of hazardous technologies. *Normal accident theory* is often contrasted to *high reliability theory*, which offers a more optimistic model for strategic planning.[11] Arguably good strategic planners would evaluate their proposed campaigns under the assumptions of both *high reliability theory* and *normal accident theory*. However, such comparisons can produce dramatically contrasting visions of the likelihood of success.

The unpredictability of complex adaptive systems, as partially captured in *normal accident theory*, underscores risks that might otherwise have been overlooked or ignored. This, however, is secondary to how much risk political leaders and military strategists consider acceptable.

Levels of Risk. As mentioned earlier, lethal autonomy is not a weapon system. It is a feature set that can be added to any weapon system. The riskiness of a specific LAW is largely a function of the destructiveness of the munition it carries.

Risk is commonly quantified as the probability of the event multiplied by its consequences. The risk posed by a weapon system rises relative to the power of the munitions the system can discharge, even when the likelihood of an adverse event occurring remains the same. Clearly, the immediate destructive impact of a machine gun pales in comparison to that of a nuclear warhead. The machine gun is inherently less risky.

It may be difficult to accurately quantify whether a specific LAW is more or less reliable than a human.

Over time, increasingly sophisticated LAWS will be deployed. Therefore it behooves the member states of CCW to not be shortsighted in their evaluation of what will be a very broad class of military applications. CCW must not appear to green light autonomous systems that can detonate weapons of mass destruction. Given the high level of risks, powerful munitions, such as autonomous ballistic missiles or autonomous submarines capable of launching nuclear warheads, must be prohibited. Deploying systems that can alter their programming is also foolhardy. This last proviso would rule out many learning systems that, for example, improve their planning capabilities.

States and military leaders may differ on the degree of unpredictability and level of risk they will accept in weapon systems. The risks posed by less powerful LAWS will, in all probability, be deemed acceptable to military strategists, particularly in comparison to similar risks posed by often unreliable humans. Nevertheless, it may be difficult to accurately quantify whether a specific LAW is more or less reliable than a human. While autonomous vehicles can be demonstrated to likely cause far fewer deaths than human drivers, similar benchmarks for accidents occurring during combat will be hard to collect and will be less than convincing. Perhaps, realistic simulated tests might demonstrate that LAWS outperformed humans in similar exercises. More importantly, the world has adjusted to accidents caused by humans. Public opinion is likely to be less forgiving of unintended wars or deaths of non-combatants caused by LAWS.

Regardless of the level of risk deemed acceptable, it is essential to recognize the degree of unpredictable risk actually posed by various autonomous weapons configurations. Empirical tools should be employed to adequately determine the risk posed by each type of LAW and whether that risk exceeds acceptable levels.

Most parties will agree that the unpredictability and therefore the risks posed by LAWS capable of dispatching high-powered munitions including nuclear weapons are unacceptable. The decision of states should not be whether any autonomous systems must be prohibited, but rather how broadly encompassing the prohibition on LAWS must be.

Mala in se

In the past, I have proposed that LAWS used for offensive purposes should be designated *mala in se*, a term coined by ancient Roman philosophers to designate an intrinsically evil activity. In just war theory and in IHL certain activities such as rape and the use of biological weapons are evil in and of themselves. Humanity's perception of evil can evolve. The ancient Romans did not consider slavery evil, but all civilized people do today. Machines that select targets and initiate lethal force are *mala in se* because they: "lack discrimination, empathy, and the capacity to make the proportional judgments necessary for weighing civilian casualties against achieving military objectives. Furthermore, delegating life and death decisions to machines is immoral because machines cannot be held responsible for their actions.[13]

Machines must not independently make choices or initiate actions that intentionally kill humans. Once this principle is in place, negotiators can move on to what will be a never-ending debate as to whether or when LAWS are extensions of human will and intention and under *meaningful human control*. With a strong moral principle in place it will be possible to condemn egregious acts.

The primary argument against this principle is the conjecture that future machines will display a capacity for discrimination and may even be more moral in their choices and actions than human soldiers.[1,2] Many in the AI and robotic community hope and believe that intelligent computational systems are becoming more than mere machines. That prospect, however, should not blind us to the opportunity to limit their destructive impact. If and when robots become ethical actors that can be held responsible for their actions, we can then begin debating whether they are no longer machines and are deserving of some form of legal personhood.

Conclusion

The short-term benefits of LAWS could be far outweighed by long-term consequences. For example, a robot arms race would not only lower the barrier to accidentally or intentionally start new wars, but could also result in a pace of combat that exceeds human response time and the reflective decision-making capabilities of commanders. Small low-cost drone swarms could turn battlefields into zones unfit for humans. The pace of warfare could escalate beyond meaningful human control. Military leaders and soldiers alike are rightfully concerned that military service will be expunged of any virtue.

In concert with the compelling legal and ethical considerations LAWS pose for IHL, unpredictability and risk concerns suggest the need for a broad prohibition. To be sure, even with a ban, bad actors will find LAWS relatively easy to assemble, camouflage, and deploy. The Great Powers, if they so desire, will find it easy to mask whether a weapon system has the capability of functioning autonomously.

The difficulties in effectively enforcing a ban are perhaps the greatest barrier to be overcome in persuading states that LAWS are unacceptable. People and states under threat perceive advanced weaponry as essential for their immediate survival. The stakes are high. No one wants to be at a disadvantage in combating a foe that violates a ban. And yet, violations of the ban against the use of biological and chemical weapons by regimes in Iraq and in Syria have not caused other states to adopt these weapons.

The power of a ban goes beyond whether it can be absolutely enforced. The development and use of biological and chemical weapons by Saddam Hussein helped justify the condemnation of the regime and the eventual invasion of Iraq. Chemical weapons use by Bashar al-Assad has been widely condemned, even if the geopolitics of the Syrian conflict have undermined effective follow-through in support of that condemnation.

The short-term benefits of LAWS could be far outweighed by long-term consequences.

A ban on LAWS is likely to be violated even more than that on biological and chemical weapons. Nevertheless, a ban makes it clear that such weapons are unacceptable and those using them are deserving of condemnation. Whenever possible that condemnation should be accompanied by political, economic, and even military measures that punish the offenders. More importantly, a ban will help slow, if not stop, an autonomous weapons arms race. But most importantly, banning LAWS will function as a moral signal that international humanitarian law (IHL) retains its normative force within the international community. Technological possibilities will not and should not succeed in pressuring the international community to sacrifice, or even compromise, the standards set by IHL.

A ban will serve to inhibit the unrestrained commercial development and sale of LAWS technology. But a preemptive ban on LAWS will not stop nor necessarily slow the roboticization of warfare. Arms manufacturers will still be able to integrate ever-advancing features into the robotic weaponry they develop. At best, it will require that a human in the loop provides a real-time authorization before a weapon system kills or destroys a target that may harbor soldiers and noncombatants alike.

Even a modest ban signals a moral victory, and will help ensure that the development of AI is pursued in a truly beneficial, robust, safe, and controllable manner. Requiring meaningful human control in the form of real-time human authorization to kill will help slow the pace of combat, but will not stop the desire for increasingly sophisticated weaponry that could potentially be used autonomously.

In spite of recent analyses suggesting that humanity has become less violent over several millennia,[9] warfare itself is an evil humanity has been unsuccessful at quelling. However, if we are to survive and evolve as a species some limits must be set on the ever more destructive and escalating weaponry technology affords. The nuclear arms race has already made clear the dangers inherent in surrendering to the inevitability of technological possibility.

Arms control will never be a simple matter. Nevertheless, we must slowly, effectively, and deliberately put a cap on inhumane weaponry and methods as we struggle to transcend the scourge of warfare.

References

1. Arkin, R. The case for banning killer robots: Counterpoint. *Commun. ACM 58*, 12 (Dec. 2015), 46–47.
2. Arkin, R. *Governing Lethal Behavior in Autonomous Systems.* CRC Press, Boca Raton, FL, 2009.
3. Asaro, P. On Banning Autonomous Lethal Systems: Human Rights, Automation and the Dehumanizing of Lethal Decision-making. Special Issue on New Technologies and Warfare. *International Review of the Red Cross 94*, 886 (Summer 2012), 687–709.
4. Carpenter, C. How do Americans feel about fully autonomous weapons? The Duck of Minerva (June 19, 2013); http://bit.ly/2mBKMnR
5. Gormley, D.M. *Missile Contagion.* Praeger Security International, 2008.
6. Gubrud, M. and Altmann, J. Compliance Measures for an Autonomous Weapons Convention, ICRAC Working Paper Series #2, International Committee for Robot Arms Control (2013); http://bit.ly/2nfOLFu
7. Peck, M. Global hawk crashes: Who's to blame? *National Defense 87*, 594 (2003); http://bit.ly/2mQJgeJ
8. Perrow, C. *Normal Accidents: Living With High-Risk Technologies.* Basic Books, New York, 1984.
9. Pinker, S. *The Better Angels of Our Nature: Why Violence Has Declined.* Penguin, 2011.
10. Roff, H. and Danks, D. Trust but Verify: The difficulty of trusting autonomous weapons systems. *Journal of Military Ethics.* (Forthcoming).
11. Sagan, S.D. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons.* Princeton University Press, Princeton, NJ, 2013.
12. Taleb, N.N. *The Black Swan: The Impact of the Highly Improbable.* Random House, 2007.
13. Wallach, W. Terminating the Terminator. *Science Progress*, 2013; http://bit.ly/2mjl2dy
14. Wallach, W. and Allen, C. Framing robot arms control. *Ethics and Information Technology 15*, 2 (2013), 125–135.

Wendell Wallach (wendell.wallach@yale.edu) is a Senior Advisor to The Hastings Center and Chairs Technology and Ethics Studies at the Yale University Interdisciplinary Center for Bioethics. His latest book is *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control.*

The author would like to express appreciation to the five anonymous reviewers, whose comments contributed significantly to help clarify the arguments made in this Viewpoint.