



ADVANCING THE DEBATE ON KILLER ROBOTS: 12 KEY ARGUMENTS FOR A PREEMPTIVE BAN ON FULLY AUTONOMOUS WEAPONS

May 2014

The development of fully autonomous weapons, also known as “killer robots,” and the proposal to ban them preemptively have sparked impassioned debate at the international and national levels. Experts—including lawyers, ethicists, military specialists, human rights advocates, and scientists—have argued about the legality and desirability of the weapons in official diplomatic meetings, at conferences around the world, in academic journals, and on the Internet. In May 2014, states parties to the Convention on Conventional Weapons (CCW) will convene in Geneva for the latest foray into the issue, a four-day experts’ meeting on what CCW states call “lethal autonomous weapons systems.” This paper seeks to advance the discussions about fully autonomous weapons by elaborating on the call for a ban and addressing head on the main arguments against such a ban.

Fully autonomous weapons, once deployed, would be able to select and fire on targets without meaningful human involvement. Although they do not yet exist, the development of precursors and military planning documents indicate that technology is moving rapidly in that direction.

Human Rights Watch and Harvard Law School’s International Human Rights Clinic (IHRC) have contributed to the global discussion with a series of papers arguing for a ban on the development, production, and use of fully autonomous weapons. In November 2012, we released *Losing Humanity: The Case against Killer Robots*, the first major civil society report on the topic. We subsequently expanded our arguments in other publications, including an analysis of the US Department of Defense’s directive on autonomous weapons, a Q&A document on fully autonomous weapons, a memorandum on the need for new law to ban these weapons, and a report on the human rights implications of the weapons.¹

¹ Human Rights Watch and Harvard Law School’s International Human Rights Clinic (IHRC), *Losing Humanity: The Case against Killer Robots*, November 2012, <http://www.hrw.org/reports/2012/11/19/losing-humanity-o>; Human Rights

Human Rights Watch and IHRC are calling on governments to:

- Work toward an international instrument prohibiting the development, production, and use of the fully autonomous weapons.
- Develop national policies on the issue, which encompass national moratoria on the development, production, and use of the fully autonomous weapons.
- Agree in November 2014 to expand CCW discussions in a more formal group of governmental experts next year, with an eye ultimately to negotiating a protocol on the weapons.

Our reports on fully autonomous weapons are part of a growing movement against the weapons. The Campaign to Stop Killer Robots, an international coalition of nongovernmental organizations (NGOs) coordinated by Human Rights Watch, has led civil society's efforts to ban the weapons. It currently has 51 member organizations from 24 countries. Other experts, including Christof Heyns, the UN special rapporteur on extrajudicial, summary or arbitrary executions, have also expressed serious concerns about the legal and moral implications of these weapons.

At the same time, critics of the campaign's position have defended the proposed technology and challenged the call for a preemptive prohibition. This paper responds directly to those critics by examining and rebutting 12 of their claims. In so doing, it seeks to add depth and nuance to the case against these weapons.

The paper is divided into 12 sections, each providing a response to a particular claim or argument that critics of a preemptive ban have made.

Watch and IHRC, "Review of the 2012 US Policy on Autonomy in Weapons Systems," April 2013, <http://www.hrw.org/news/2013/04/15/review-2012-us-policy-autonomy-weapons-systems>; Human Rights Watch and IHRC, "Q&A on Fully Autonomous Weapons," October 2013, <http://www.hrw.org/news/2013/10/21/qa-fully-autonomous-weapons>; Human Rights Watch and IHRC, "The Need for New Law to Ban Fully Autonomous Weapons," November 2013, <http://www.hrw.org/news/2013/11/13/need-new-law-ban-fully-autonomous-weapons>; Human Rights Watch and IHRC, *Shaking the Foundations: The Human Rights Implications of Killer Robots*, May 2014, <http://hrw.org/node/125251>.

Claim #1: A new treaty is unnecessary because existing international humanitarian law is adequate.

Response: A new treaty would help clarify existing international humanitarian law and would address development and production of fully autonomous weapons in addition to their use.

Analysis: Critics of a new treaty on fully autonomous weapons often assert that existing international humanitarian law is sufficient to deal with the dangers posed by the weapons. They argue that any problematic use of fully autonomous weapons would already be unlawful because it would violate current international humanitarian law. According to two authors, “The question for the legal community [would be] whether autonomous weapon systems comply with the legal norms that States have put in place.”² Recognizing that the weapons raise new concerns, another author notes that “as cases and mistakes arise, the lawyers and injured parties will have to creatively navigate the network of legal mechanisms [available in international law].”³ Yet he too concludes that a new legal instrument would be unnecessary.

Existing international humanitarian law, however, was not intended to and cannot adequately address the issues raised by fully autonomous weapons. International humanitarian law should be supplemented with new law designed to deal with the unique challenges of this revolutionary type of weapon.

A new international treaty would clarify states’ obligations and make explicit the requirements for compliance. It would minimize questions about legality by standardizing rules across countries and reducing the need for case-by-case determinations. Greater legal clarity would lead to more effective enforcement because countries would better understand the rules. A ban convention would make the illegality of fully autonomous weapons clear even for countries that do not conduct legal reviews of new or modified weapons. Finally, many states that do not join the new treaty would still be apt to abide by its ban because of the stigma associated with the weapons.

² Michael N. Schmitt and Jeffrey S. Thurnher, “‘Out of the Loop’: Autonomous Weapon Systems and the Law of Armed Conflict,” *Harvard National Security Journal*, vol. 4 (2013), p. 232.

³ Benjamin Kastan, “Autonomous Weapons Systems: A Coming Legal ‘Singularity’?” *University of Illinois Journal of Law, Technology, and Policy* (Spring 2013), p. 45.

A treaty dedicated to fully autonomous weapons could also address aspects of proliferation not covered under traditional international humanitarian law, which focuses on the use of weapons in war. In particular, such an instrument could prohibit development and production. Eliminating these activities would prevent the spread of fully autonomous weapons, including to states or non-state actors with little regard for international humanitarian law or limited ability to enforce compliance. In addition, it would help avert an arms race by stopping development before it went too far.

Finally, new law could address concerns about an accountability gap under existing international humanitarian law (see more detailed discussion of accountability under claim #6 below). A treaty could establish a ban under any circumstances and specify that anyone violating that rule would be held responsible for the fully autonomous weapon's actions.

While international humanitarian law already sets limits on problematic weapons and their use, responsible governments have in the past found it necessary to supplement existing legal frameworks for weapons that by their nature pose significant humanitarian threats. Treaties dedicated to specific weapons types exist for cluster munitions, antipersonnel mines, blinding lasers, chemical weapons, and biological weapons. Fully autonomous weapons have the potential to raise a comparable or even higher level of humanitarian concern and thus should be the subject of similar supplementary international law.

Claim #2: Continued developments in artificial intelligence might make it possible for fully autonomous weapons to comply with the principles of distinction and proportionality, at least in certain circumstances.

Response: It is likely that fully autonomous weapons would never be capable of reliably complying with the principles of distinction and proportionality.

Analysis: Critics argue that advocates of a ban often “fail to take account of likely developments in autonomous weapon systems technology.”⁴ According to these critics, not only has military technology “advanced well beyond simply being able to spot an individual or object,” but improvements in artificial intelligence will probably also

⁴ Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 234.

continue.⁵ Thus, while recognizing the existence of “outstanding issues” and “daunting problems,”⁶ critics are content with the belief that solutions are “theoretically achievable.”⁷ Human Rights Watch and IHRC do not share this optimism and see problems in assuming that such weapons can ever conform to the international humanitarian law requirements of distinction and proportionality.

Distinction

Fully autonomous weapons would face great, if not insurmountable, difficulties in reliably distinguishing between lawful and unlawful targets as required by international humanitarian law. Although progress is likely in deep sensory and processing capabilities for these weapons, replicating the elements of human judgment is far more difficult to achieve. The weapons would lack human qualities that facilitate making such determinations, particularly on contemporary battlefields where combatants often seek to conceal their identities. Distinguishing an active combatant from a civilian or injured or surrendering soldier requires more than deep sensory and processing capabilities. It also depends on the qualitative ability to gauge human intention, which involves interpreting subtle, context-dependent clues, such as tone of voice, facial expressions, or body language. Humans possess the unique capacity to identify with other human beings and are thus better equipped to understand the nuances of unforeseen behavior in ways in which machines—which must be programmed in advance—simply cannot.

Proportionality

The obstacles presented by the principle of distinction are compounded when it comes to proportionality, which requires the delicate balancing of two factors: expected civilian harm and anticipated military advantage. This evaluation takes place not only in anticipation of an overall military battle plan but also during actual military operations, when decisions must be made about the course or cessation of any particular attack. When it comes to expected civilian harm, one critic has concluded that there “is no question that autonomous weapon systems could be programmed ... to determine the likelihood of harm to civilians in the target area.”⁸ Similarly, while

⁵ Michael N. Schmitt, “Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics,” *Harvard National Security Journal Features* online (2013). <http://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/> (accessed May 8, 2014), p. 11.

⁶ Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Raton, FL: CRC Press, 2009), pp. 126, 211.

⁷ Schmitt, “Autonomous Weapon Systems and International Humanitarian Law,” *Harvard National Security Journal Features*, p. 17 (discussing, in particular, whether autonomous weapons could be programmed to adequately “compute doubt”).

⁸ *Ibid.*, p. 20.

acknowledging that “it is unlikely in the near future that ... ‘machines’ will be programmable to perform robust assessments of a strike’s likely military advantage,” the same critic has written that “military advantage algorithms could *in theory* be programmed into autonomous weapon systems.”⁹

There are a number of reasons seriously to doubt each of these conclusions. As already discussed, it is highly questionable whether a fully autonomous weapon could ever reliably distinguish legitimate from illegitimate targets. This doubt is enhanced where it is not only the legitimacy of the target that is in question, but also the expected civilian harm—a calculation that requires determining the status of and impact on all the possible entities and objects surrounding the target that might be affected by an attack.

When it comes to predicting anticipated military advantage, even critics admit that “doing so will be challenging [for a machine] because military advantage determinations are always contextual.”¹⁰ More specifically, because military advantage must be determined on a “case-by-case” basis, it is unclear how a programmer could account, in advance, for the infinite variety of unexpected contingencies that may arise in a deployment.¹¹

Even if the elements of military advantage and expected civilian harm could be adequately quantified by a fully autonomous weapon system, such a system would be unlikely to be able qualitatively to balance them. The generally accepted standard for assessing proportionality is whether a “reasonable military commander” would have launched a particular attack.¹²

In weighing the proportionality of an attack by a fully autonomous weapon, the appropriate question would be whether the weapon system made a reasonable targeting determination at the time of its strike. Some have suggested the pertinent question with fully autonomous weapons is whether a human commander acted reasonably in deploying it ahead of the strike.¹³ The proportionality of any particular

⁹ Ibid. (emphasis added).

¹⁰ Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 255.

¹¹ For a discussion of the case-by-case nature of proportionality, see *ibid.*, p. 256 (asserting that “the military advantage element of the proportionality rule generally necessitates case-by-case determinations”).

¹² “Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia,” <http://www.icty.org/sid/10052> (accessed May 8, 2014), para. 50.

¹³ See Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 280 (“Human operators, not machines or software, will ... be making the subjective determinations required under the law of armed conflict, such as those involved in proportionality or precautions in attack calculations. Although the subjective decisions may

attack cannot be ensured at the time of deployment, however, and the decision to deploy is not an equivalent determination to the decision to attack. A commander considering whether to deploy a fully autonomous weapon would need to rely on the programmer's and manufacturer's predictions of how the weapon would perform in potentially shifting or unforeseeable conditions. No matter how much care was taken, a programmer or manufacturer would be unlikely to anticipate all conditions that would affect the machine's performance, which would exacerbate the challenge the commander would face in determining proportionality. In addition, advance programming and testing can never replace human control in the course of operations, and the human judgment to deploy a weapon is not the same as that to mount, continue, alter, or terminate an attack. Yet at the moment of making a determination to attack, a fully autonomous weapon would neither be under the control of a human being exercising his or her own judgment nor able to exercise genuine human judgment itself.

It would be difficult to create machines that could meet the reasonable military commander standard and be expected to act "reasonably" when making determinations to attack in unforeseen or changeable circumstances. According to the Max Planck Encyclopedia of International Law, "[t]he concept of reasonableness exhibits an important link with *human* reason," and it is "generally perceived as opening the door to several *ethical* or *moral*, rather than legal, considerations."¹⁴ Two critics of the proposed ban treaty have noted that "[p]roportionality ... is partly a technical issue of designing systems capable of measuring predicted civilian harm, but also partly an ethical issue of attaching weights to the variables at stake."¹⁵ Many people would object to the idea that machines could or should be making ethical or moral determinations, and yet this is precisely what the reasonable military commander standard requires. Moreover, reasonableness eludes "objective definition" and depends on the situation.¹⁶ If humans cannot know in advance what would be reasonable in every given situation (because making determinations is context-

sometimes have to be made earlier in the targeting cycle than has traditionally been the case, this neither precludes the lawfulness of the decisions, nor represents an impediment to the lawful deployment of the systems.").

¹⁴ Olivier Corten, "Reasonableness in International Law," *Max Planck Encyclopedia of Public International Law*, updated May 2006, <http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1679?rskey=U1bcaw&result=10&prd=EPIL> (accessed February 21, 2014), para. 1 (emphasis added).

¹⁵ Kenneth Anderson and Matthew Waxman, "Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can," *Jean Perkins Task Force on National Security and Law*, http://media.hoover.org/sites/default/files/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf (accessed May 8, 2014), p. 23.

¹⁶ Corten, "Reasonableness in International Law," *Max Planck Encyclopedia of Public International Law*, para. 1.

specific), it is unrealistic to expect programmers to program machines to act reasonably in inherently unforeseeable situations.

While proportionality analyses allow for a “fairly broad margin of judgment,” the sort of judgment required in deciding how to weigh civilian harm and military advantage in unanticipated situations would be difficult to replicate in machines.¹⁷ As Christof Heyns, the UN special rapporteur on extrajudicial, summary or arbitrary executions, has explained, assessing proportionality requires “distinctively human judgement.”¹⁸ According to the International Committee of the Red Cross (ICRC), judgments about whether a particular attack is proportionate “must above all be a question of *common sense* and *good faith*,” characteristics that many would agree machines cannot possess, however thorough their programming.¹⁹

Claim #3: Fully autonomous weapons should not be treated as per se unlawful because they could be used lawfully in some circumstances.

Response: Narrowly constructed hypothetical cases in which fully autonomous weapons could lawfully be used should not be employed to legitimize the weapons or stand in the way of a ban because the cases do not alter the underlying concerns about the use of such weapons.

Analysis: Critics argue that fully autonomous weapons would not be unlawful per se because there are some potential uses, no matter how limited or unlikely, where they would be both militarily valuable and capable of conforming to the requirements of international humanitarian law. One critic, for example, notes that “[n]ot every battlespace contains civilians.”²⁰ Other critics maintain fully autonomous weapons

¹⁷ International Committee of the Red Cross, Commentary of 1987 on Article 57 of the Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), adopted 8 June 1977, <http://www.icrc.org/ihl.nsf/INTRO/470> (accessed May 8, 2014), para. 2210.

¹⁸ UN Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, Lethal Autonomous Robotics, A/HRC/23/47, April 9, 2013, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (accessed May 8, 2014), p. 14. See also Human Rights Watch and IHRC, *Losing Humanity*, pp. 32-34 (noting that because the proportionality test is a subjective one, it requires human judgment, “rather than the automatic decision making characteristic of a computer”).

¹⁹ International Committee of the Red Cross, Commentary of 1987 on Article 57 of Protocol I, para. 2208 (emphasis added).

²⁰ Schmitt, “Autonomous Weapon Systems and International Humanitarian Law,” *Harvard National Security Journal Features*, p. 11.

could be used lawfully under “limited circumstances,” such as in attacks on “nuclear-tipped mobile missile launchers, where millions of lives were at stake.”²¹

One can almost always construct a hypothetical situation where use of a widely condemned weapon could arguably comply with the general rules of international humanitarian law. Before adoption of the 2008 Convention on Cluster Munitions, proponents of cluster munitions often maintained that the weapons could be lawfully launched on a military target alone in an otherwise unpopulated desert. It is extremely difficult, however, to restrict use of weapons to narrowly constructed scenarios, as exemplified by the widespread use of cluster munitions in populated areas. Such theoretical possibilities should not be used to legitimize weapons, including fully autonomous ones, that pose significant humanitarian risks when used in less exceptional situations. The small chance of lawful use in limited circumstances should also not stand in the way of an international prohibition.

Claim #4: Due to the legal requirement to take precautions in attacks, militaries would only use fully autonomous weapons if they were the most humanitarian option among equally effective weapons.

Response: Users of fully autonomous weapons would likely disregard the obligation to take precautions in some circumstances.

Analysis: Some critics object to a ban because they argue that international humanitarian law’s rule on taking precautions in attacks would limit the use of fully autonomous weapons to situations in which they were the most humanitarian alternative. This rule requires militaries to “take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects.”²² Under this provision, critics contend, fully autonomous weapons would only be used if they were the option that would have the least civilian impact among weapons that would achieve the “desired military objective.”²³ One author explains that “the only

²¹ Paul Scharre, “Reflections on the Chatham House Autonomy Conference,” *Lawfare* blog, <http://www.lawfareblog.com/2014/03/guest-post-reflections-on-the-chatham-house-autonomy-conference/> (accessed April 20, 2014).

²² Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), adopted June 8, 1977, 1125 U.N.T.S. 3, entered into force December 7, 1978, art. 57(2)(a)(ii).

²³ Schmitt, “Autonomous Weapon Systems and International Humanitarian Law,” *Harvard National Security Journal Features*, p. 24.

situation in which an autonomous weapon system can lawfully be employed is when its use will realize military objectives that cannot be attained by other available systems that would cause less collateral damage.”²⁴

The critics’ position, however, depends on fully autonomous weapons being used only in compliance with international humanitarian law. Some parties might have little respect for this body of law, while others who generally comply with it might be tempted to use their fully autonomous weapons in certain dire circumstances (see further discussion under claim #10). The critics ignore these scenarios and the magnitude of the consequences if militaries used the weapons when they were not the most humanitarian option.

A ban would strive to prevent such situations by preempting the creation and proliferation of fully autonomous weapons as well as their use. It would also strongly stigmatize the weapons, putting political pressure on all parties not to use them.

Claim #5: Fully autonomous weapons would not be negatively influenced by human emotions when making determinations to use lethal force.

Response: Fully autonomous weapons would lack emotions, including compassion and a resistance to killing, that can protect civilians and soldiers.

Analysis: Critics argue that fully autonomous weapons’ lack of human emotions could have military and humanitarian benefits. The weapons would be immune from factors, such as fear, anger, pain, and hunger, that can cloud judgment, distract humans from their military missions, or lead to attacks on civilians.²⁵ While such observations have some merit, the role in warfare of other human emotions can in fact advance humanitarian protection in armed conflict.

Humans possess empathy and compassion and are generally reluctant to take the life of another human. A retired US Army Ranger who has done extensive research on killing during war has found that “there is within man an intense resistance to killing

²⁴ Ibid.

²⁵ Ronald C. Arkin, “Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,” Technical Report GIT-GVU-07-11, <http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf> (accessed May 8, 2014), pp. 6-7.

their fellow man. A resistance so strong that, in many circumstances, soldiers on the battlefield will die before they can overcome it.”²⁶ Another author writes,

One of the greatest restraints for the cruelty in war has always been the natural inhibition of humans not to kill or hurt fellow human beings. The natural inhibition is, in fact, so strong that most people would rather die than kill somebody.²⁷

Studies of soldiers’ conduct in past conflicts provide evidence to support these conclusions.²⁸ Human emotions are thus an important inhibitor to killing people unlawfully or needlessly.

Studies have focused largely on troops’ reluctance to kill enemy combatants, but it is reasonable to assume that soldiers feel even greater reluctance to kill the bystanders of war, including civilians or those *hors de combat*, such as surrendering or wounded soldiers. Fully autonomous weapons, unlike humans, would lack such emotional and moral inhibitions, which, while not required by international law, help protect individuals who are not lawful targets in an armed conflict. One expert writes, “Taking away the inhibition to kill by using robots for the job could weaken the most powerful psychological and ethical restraint in war. War would be inhumanely efficient and would no longer be constrained by the natural urge of soldiers not to kill.”²⁹

Due to their lack of emotion, fully autonomous weapons could be the perfect tools for leaders who seek to oppress their own people or to attack civilians in enemy countries. Even the most hardened troops can eventually turn on their leader if ordered to fire on their own people or to commit war crimes. An abusive leader who can resort to fully

²⁶ Lt. Col. Dave Grossman, *On Killing: The Psychological Cost of Learning to Kill in War and Society* (New York: Little, Brown and Company, 1995), p. 4.

²⁷ Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Farnham: Ashgate Publishing Limited, 2009), p. 130.

²⁸ For example, based on interviews with thousands of US soldiers in World War II, US Army Brig. Gen. S.L.A. Marshall found that usually only 15 to 20 percent of troops would fire at the enemy. These numbers were due to an innate hesitancy to kill, not to fear or cowardice, because “[t]hose who would not fire did not run or hide (and in many cases they were willing to risk great danger to rescue comrades, get ammunition, or run messages).” S.L.A. Marshall, *Men against Fire: The Problem of Battle Command in Future War* (New York: William Morrow & Company, 1947), p. 54; Grossman, *On Killing*, p. 4. Other researchers have documented how troops avoided killing by repeatedly loading their guns without firing or by shooting over the enemies’ heads. For discussion of troops in US Civil War repeatedly loading their rifles, see Grossman, *On Killing*, pp. 18-28. For discussion of Ardant du Picq’s study on nineteenth-century French troops firing in the air, see Grossman, *On Killing*, pp 9-10. See also Grossman, *On Killing*, pp. 16-17 (discussing a 1986 study by British Defense Operational Analysis Establishment of 100 “nineteenth- and twentieth-century battles and test trials”).

²⁹ Krishnan, *Killer Robots*, p. 130.

autonomous weapons would be free of the fear that armed forces would resist being deployed against certain targets.

For all the reasons outlined above, rather than being understood as irrational influences and obstacles to reason, emotions should instead be viewed as central to restraint in war.

Claim #6: Existing international humanitarian law can adequately address accountability concerns arising out of the use of fully autonomous weapons.

Response: Insurmountable legal and practical obstacles would likely interfere with holding someone accountable for unforeseeable, unlawful acts committed by a fully autonomous weapon.

Analysis: Critics argue that the “mere fact that a human might not be in control of a particular engagement does not mean that no human is responsible for the actions of the autonomous weapon system.”³⁰ According to these critics, “[a] human must decide how to program the system and when to launch it.”³¹ Thus, a programmer or manufacturer could be held accountable for intentionally creating a robot that would commit war crimes, and the person deploying the robot “would be accountable for those war crimes if he or she knew or should have known that the autonomous weapon system had been so programmed and did nothing to stop its use.”³²

Certainly a commander could be held responsible for intentionally using a fully autonomous weapon that was clearly unsuited for the environment in which it was deployed.³³ Furthermore, a programmer or manufacturer could be held liable for intentionally producing a fully autonomous weapon that would commit war crimes. These scenarios, however, fail to capture perhaps more likely situations in which the commander, programmer, or manufacturer did not know a robot would commit an

³⁰ Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 277.

³¹ *Ibid.*

³² *Ibid.*

³³ *Ibid.* (“Hopefully, autonomous weapon systems will never be programmed to commit war crimes. Much more likely would be a case in which a system that has not been so programmed is nevertheless used in a manner that constitutes such crimes. For example, the operator of an autonomous weapon system that cannot distinguish civilians from combatants who employs the system in an area where the two are intermingled has committed the war crime of indiscriminate attack.”).

illegal act, but the robot nonetheless unexpectedly did so. In such cases, there would be no human to hold directly responsible for the decision to attack, and indirect liability would be difficult to achieve.

Due to the autonomous nature of their technology, fully autonomous weapons could act unforeseeably in ways that would cause unlawful harm. One type of war crime is the act of willfully “launching an indiscriminate attack affecting the civilian population or civilian objects *in the knowledge* that such attack will cause excessive loss of life, injury to civilians or damage to civilian objects.”³⁴ Given the technological limitations discussed under claim #2, a fully autonomous weapon, created and deployed to comply with international humanitarian law, might nonetheless launch an attack whose military advantage was so obviously outweighed by its civilian costs that any reasonable human in its position would have known that the attack was disproportionate. Similarly, a fully autonomous weapon might select and fire upon a civilian target mistaking it for a military target, even when a reasonable human would have known that the object of the attack was civilian.

Significant obstacles to holding anyone accountable would exist for both of these situations. Robots themselves could not be punished for committing war crimes as they lack the capacity to feel pain or other emotions associated with punishment. Command responsibility holds military commanders responsible for subordinates’ actions if they knew or should have known their subordinates committed or were going to commit a crime and failed to prevent the crime or punish the subordinates.³⁵ In the circumstances described above, the commander could not foresee and thus not prevent the violations in question, and he or she could not punish the robot after the fact.

An alternative option would be to try to hold the programmer or manufacturer civilly liable for the unanticipated acts of a fully autonomous weapon. Tort law offers an approach other than prosecution, but it too would likely fail to ensure accountability. In the United States, for example, defense contractors are generally not found liable for harm caused by their products.³⁶ Even without a legal gap, there are policy and

³⁴ Protocol I, art. 85 (emphasis added).

³⁵ See, for example, Rome Statute of the International Criminal Court (Rome Statute), U.N. Doc. A/CONF.183/9, July 17, 1998, entered into force July 1, 2002, art. 28; Statute of the International Criminal Tribunal for the former Yugoslavia (ICTY Statute), S.C. Res. 827, U.N. Doc. S/RES/827 (1993), as amended, art. 7(3).

³⁶ Under the Federal Torts Claims Act, the government waives its immunity from civil suits in certain situations. The Supreme Court has applied this rule to contractors hired by the government. The waiver, however, is subject to the discretionary function exception and the combatant activities exception, which would block most suits against

practical problems with holding programmers and manufacturers accountable. Such liability could be unfair since even programmers and manufacturers might be unable to foresee the harm their fully autonomous weapons could cause in various situations.³⁷ In addition, civil suits are generally brought by victims and, especially in cases of armed conflict, it is unrealistic to think all victims would have the resources or adequate access to obtain justice. This practical limitation is significant because civil litigation against those who program, manufacture, or use such robots would be a more likely avenue of redress than prosecution.³⁸

The use of fully autonomous weapons would thus lead to the creation of a potentially insurmountable accountability gap. The lack of criminal or civil consequences would interfere with deterrence. A failure to punish would leave victims and their relatives without the satisfaction that someone paid for the suffering they experienced.

Claim #7: The Martens Clause does not restrict the use of fully autonomous weapons.

Response: Because existing law does not specifically address the unique issues raised by fully autonomous weapons, the Martens Clause mandates that the “principles of humanity” and “dictates of public conscience” be factored into an analysis of their legality. Both of these standards weigh in favor of a ban on this kind of technology.

Analysis: Some critics dismiss the value of the Martens Clause in determining the legality of fully autonomous weapons. As it appears in Additional Protocol I, the Martens Clause mandates that:

programmers and manufacturers of fully autonomous weapons. Andrew Finkelman, “Suing the Hired Guns: An Analysis of Two Federal Defenses to Tort Lawsuits against Military Contractors,” *Brooklyn Journal of International Law*, vol. 34 (2009), p. 405. (“There are two important exceptions at the heart of the military contractor’s efforts to invoke federal sovereign immunity. The first exception is the ‘discretionary-function’ exception, which preserves federal sovereign immunity in suits arising from ‘the exercise or performance or the failure to exercise or perform a discretionary function’ by the federal government or its employees.... The second exception is the ‘combatant-activities’ exception, which excepts suits against the federal government ‘arising out of the combatant activities of the military or naval forces, or the Coast Guard, during time of war.’”).

³⁷ One commentator said assigning them responsibility would be like “holding parents accountable for the actions of their children after they have left their care.” Robert Sparrow, “Killer Robots,” *Journal of Applied Philosophy*, vol. 24 (2007), p. 70. See generally Andreas Matthias, “The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata,” *Ethics and Information Technology*, vol. 6 (2004), pp. 175-183.

³⁸ UN Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, *Lethal Autonomous Robotics*, p. 15.

[i]n cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.³⁹

Critics argue that the Martens Clause “does not act as an overarching principle that must be considered in every case,” but is, rather, merely “a failsafe mechanism meant to address lacunae in the law.”⁴⁰ They contend that because gaps in the law are rare, “the likelihood that” autonomous weapon systems would violate the Martens Clause but not “applicable treaty and customary law” is therefore “exceptionally low.”⁴¹ The lack of specific law on fully autonomous weapons, however, means that the Martens Clause does apply, and the weapons raise serious concerns under the provision.

The key question in determining the relevance of the Martens Clause to fully autonomous weapons is the extent to which such weapons are “covered” by existing treaty law. As the US Military Tribunal at Nuremberg explained, the Martens Clause makes “the usages established among civilized nations, the laws of humanity and the dictates of public conscience into the legal yardstick to be applied if and when the *specific* provisions of [existing law] do not cover *specific* cases occurring in warfare.”⁴² The International Court of Justice asserted that the clause’s “continuing existence and applicability is not to be doubted” and that it has “proved to be an effective means of addressing the rapid evolution of military technology.”⁴³ Fully autonomous weapon systems are rapidly evolving forms of technology that are at best only generally covered by existing law.⁴⁴

³⁹ Protocol 1, art 1(2).

⁴⁰ Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 275.

⁴¹ *Ibid.*, p. 276.

⁴² *In re Krupp*, US Military Tribunal Nuremberg, judgment of July 31, 1948, in *Trials of War Criminals Before the Nuremberg Military Tribunals*, vol. IX, p. 1340 (emphasis added).

⁴³ International Court of Justice, *Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons*, July 8, 1996, <http://www.icj-cij.org/docket/files/95/7495.pdf> (accessed May 8, 2014), para. 78.

⁴⁴ Some critics argue international humanitarian law would adequately cover autonomous weapon systems, but the most relevant rules are general ones, such as those of distinction and proportionality discussed above under claim #2. While critics also emphasize the applicability of disarmament treaties on antipersonnel landmines, cluster munitions, and incendiary weapons, these instruments do not provide specific law on fully autonomous weapons. They would only govern the payload of narrow group of such weapons and would not address the challenging issues unique to autonomous systems. To date, there is no specific law dedicated to fully autonomous weapons. For critics’ view, see Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 276.

The plain language of the Martens Clause elevates the “principles of humanity” and the “dictates of public conscience” to independent legal standards against which new forms of military technology should be evaluated.⁴⁵ On this basis, any weapon conflicting with either of these standards is therefore arguably unlawful. At a minimum, however, the dictates of public conscience and principles of humanity can “serve as fundamental guidance in the interpretation of international customary or treaty rules.”⁴⁶ Thus, “[i]n case of doubt, international rules, in particular rules belonging to humanitarian law, must be construed so as to be consonant with general standards of humanity and the demands of public conscience.”⁴⁷ Given the serious doubts about the ability of fully autonomous weapons to conform to the requirements of the law, as discussed under claim #2, the standards of the Martens Clause should at the very least be taken into account when evaluating them.

Fully autonomous weapons raise serious concerns under the principles of humanity and dictates of public conscience. The ICRC has described the principles of humanity as requiring compassion and the ability to protect.⁴⁸ The challenges the weapons would face in meeting international humanitarian law suggest they could not adequately protect civilians. As discussed above under claim #5, robots would also lack human emotions, including compassion. Public opinion can play a role in revealing and shaping public conscience, and many people find the prospect of delegating life-and-death decisions to machines shocking and unacceptable. A 2013 national representative survey of 1,000 Americans found that, of those with a view, two-thirds came out against fully autonomous weapons: 68 percent opposed the move toward these weapons (48 percent strongly), while 32 percent favored their

⁴⁵ See, for example, *In re Krupp*, US Military Tribunal Nuremberg, p. 1340 (asserting that the Martens Clause “is much more than a pious declaration”). See also Antonio Cassese, “The Martens Clause: Half a Loaf or Simply Pie in the Sky?” *European Journal of International Law*, vol. 11, no. 1 (2000), p. 210 (asserting that “a majority of states appearing before the [International Court of Justice]” with regards to the *Nuclear Weapons Advisory Opinion* “suggested—either implicitly or in a convoluted way—the expansion of the scope of the clause so as to upgrade it to the rank of a norm establishing new sources of law”); International Committee of the Red Cross, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977* (2006), <http://www.icrc.org/eng/resources/documents/publication/p0902.htm> (accessed May 8, 2014), p. 17 (stating that “[a] weapon which is not covered by existing rules of international humanitarian law would be considered contrary to the Martens [C]ause if it is determined per se to contravene the principles of humanity or the dictates of public conscience”).

⁴⁶ Cassese, “The Martens Clause,” *European Journal of International Law*, p. 212.

⁴⁷ *Ibid.* See also Jochen von Bernstorff, “Martens Clause,” *Max Planck Encyclopedia of Public International Law*, updated December 2009, <http://opil.ouplaw.com.ezp-prod1.hul.harvard.edu/view/10.1093/law:epil/9780199231690/law-9780199231690-e327?rskey=QVxFkp&result=1&prd=EPL> (accessed May 8, 2014), para. 13 (“A second reading sees the clause as an interpretative device according to which, in case of doubt, rules of international humanitarian law should be interpreted according to ‘principles of humanity’ and ‘dictates of public conscience’.”).

⁴⁸ International Committee of the Red Cross, *The Fundamental Principles of the Red Cross and Red Crescent*, ICRC Publication ref. 0513 (1996), http://www.icrc.org/eng/assets/files/other/icrc_002_0513.pdf (accessed November 3, 2013), p. 2.

development.⁴⁹ Interestingly, active duty military personnel were among the strongest objectors—73 percent expressed opposition to fully autonomous weapons. These kinds of reactions indicate that fully autonomous weapons could contravene the Martens Clause.

Concerns about weapons' compliance with the principles in the Martens Clause have justified new weapons treaties in the past. For example, the Martens Clause heavily influenced the discussions and debates preceding the development of CCW Protocol IV on Blinding Lasers, which bans the transfer and use of laser weapons whose sole or partial purpose is to cause permanent blindness.⁵⁰ The Martens Clause was invoked not only by civil society in its reports on the matter, but also by experts participating in a series of ICRC meetings on the subject.⁵¹ They largely agreed that “[blinding lasers] would run counter to the requirements of established custom, humanity, and public conscience.”⁵² It appears that a shared visceral reaction against blinding weapons ultimately tipped the scales toward a prohibition, even without consensus that such weapons were unlawful under the core principles of international humanitarian law.⁵³ The Blinding Lasers Protocol set an international precedent for preemptively banning weapons based at least in part on the Martens Clause.⁵⁴ Invoking the clause in the context of fully autonomous weapons would be equally appropriate.

⁴⁹ Charli Carpenter, “US Public Opinion on Autonomous Weapons,” June 19, 2013, http://www.whiteoliphant.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf (accessed May 8, 2014). Many who responded “not sure” preferred a precautionary approach “in the absence of information.” Charli Carpenter, “How Do Americans Feel about Fully Autonomous Weapons?” *Duck of Minerva*, June 19, 2013, <http://www.whiteoliphant.com/duckofminerva/2013/06/how-do-americans-feel-about-fully-autonomous-weapons.html> (accessed May 8, 2014). These figures are based on a nationally representative online poll of 1,000 Americans conducted by Yougov.com. Respondents were an invited group of Internet users (YouGov Panel) matched and weighted on gender, age, race, income, region, education, party identification, voter registration, ideology, political interest, and military status. The margin of error for results is +/- 3.6%. A discussion of the sampling methods, limitations, and accuracy can be found at <http://yougov.co.uk/publicopinion/methodology/>.

⁵⁰ David Akerson, “The Illegality of Offensive Lethal Autonomy,” in *International Humanitarian Law and the Changing Technology of War*, ed. Dan Saxon (Leiden: Martinus Nijhoff, 2013), pp. 92-93; Convention on Conventional Weapons, Protocol IV on Blinding Lasers (CCW Protocol IV), adopted October 13, 1995, entered into force July 30, 1998, art. 1.

⁵¹ See, for example, Human Rights Watch, *Blinding Laser Weapons: The Need to Ban a Cruel and Inhumane Weapon*, vol. 7, no. 1 (1995), http://www.hrw.org/reports/1995/General1.htm#P583_118685.

⁵² According to the ICRC report, “some experts expressed either personal repugnance for lasers or the belief that their countries' civilian population would find the use of blinding as a method of warfare horrific.” International Committee of the Red Cross, *Blinding Weapons: Reports of the Meetings of Experts Convened by the International Committee of the Red Cross on Battlefield Laser Weapons, 1989-1991* (Geneva: International Committee of the Red Cross, 1993), pp. 344-46. Others doubted their ability to field such weapons, notwithstanding possible military utility, because of public opinion. *Ibid.*, p. 341.

⁵³ This visceral reaction is suggested by the comments of the participating experts in the ICRC meetings. Examples include the statement of one participant that he would be unable to introduce blinding weapons in his country “because public opinion would be repulsed at the idea.” Another participant described it as “indisputable that deliberately blinding on the battlefield would be socially unacceptable.” *Ibid.*, p. 345.

⁵⁴ Akerson, “The Illegality of Offensive Lethal Autonomy,” p. 96.

Claim #8: A ban on fully autonomous weapons is premature given the possibility of a technological fix.

Response: These highly problematic weapons should be preemptively banned to prevent serious humanitarian harm before it is too late and to accord with the precautionary principle.

Analysis: Critics contend that a preemptive ban on the development, production, and use of fully autonomous weapons is premature. They argue that:

[r]esearch into the possibilities of autonomous machine decision-making, not just in weapons but across many human activities, is only a couple of decades old.... We should not rule out in advance possibilities of positive technological outcomes—including the development of technologies of war that might reduce risks to civilians by making targeting more precise and firing decisions more controlled.⁵⁵

This position depends in part on one's faith that technology could address the legal challenges raised by fully autonomous weapons, which, as explained above under claim #2, Human Rights Watch and IHRC question. At the same time, it ignores other problems associated with these weapons that are not related to technology, notably the potential for an arms race, an accountability gap, and moral objections, which are discussed under claims #1, 6, and 9.

Given the host of concerns about fully autonomous weapons, they should be preemptively banned before it becomes too late to change course. It is difficult to stop technology once large-scale investments have been made. The temptation to use technology already developed and incorporated into military arsenals would be great, and many countries would be reluctant to give it up, especially if their competitors were deploying it.

In addition, if ongoing development were permitted, militaries might deploy fully autonomous weapons in complex circumstances before artificial intelligence could handle them. Only after the weapons faced unanticipated situations that they were not programmed to address could the technology be modified to resolve those issues.

⁵⁵ Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems," *Jean Perkins Task Force on National Security and Law*, p. 15.

During that period, the weapon would be likely to mishandle such situations potentially causing great civilian harm.

The prevalence of humanitarian concerns and the uncertainty regarding technology make it appropriate to invoke the precautionary principle, a principle of international law. The 1992 Rio Declaration states, “Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation.”⁵⁶ While the Rio Declaration applies the precautionary principle to environmental protection, the principle can be adapted to other situations.

Fully autonomous weapons implicate the three essential elements of the precautionary principle—threat of serious or irreversible damage, scientific uncertainty, and the availability of cost-effective measures to prevent harm. The development, production, and use of fully autonomous weapons present a threat to civilians that would be both serious and irreversible, as the technology would revolutionize armed conflict and would be difficult to eliminate once developed and employed. Scientific uncertainty characterizes the debate over these weapons. Defenders argue there is no proof that a technological fix could not solve the problem, but there is an equal lack of proof that a technological fix would work. Finally, while treaty negotiations and implementation would carry costs, these expenses are small compared to the significant harm they might prevent.

There is precedent for a preemptive ban on a class of weapons. As discussed under claim #7 above, in 1995 governments agreed to a ban on blinding lasers before the weapons had started to be deployed out of concerns for the humanitarian harm the weapons would cause.⁵⁷ While an international, preemptive ban should be the end goal, a first step could be national moratoria on fully autonomous weapons, such as those proposed by Christof Heyns, the UN special rapporteur on extrajudicial, summary or arbitrary executions, in his May 2013 report to the Human Rights Council.⁵⁸

⁵⁶ Rio Declaration on Environment and Development, adopted June 14, 1992, U.N. Doc. A/CONF.151/26 (vol. 1), 31 ILM 874, 1992, principle 15. The Rio Declaration was a product of the 1992 United Nations Conference on Environment and Development. The UN Conference addressed growing concern over risks of environmental degradation and was attended by representatives from 172 nations. UN Conference on Environment and Development (1992), <http://www.un.org/geninfo/bp/enviro.html> (accessed May 8, 2014).

⁵⁷ CCW Protocol IV. See also International Committee of the Red Cross, “Ban on Blinding Laser Weapons Now in Force,” news release 98/31, July 30, 1998, <http://www.icrc.org/eng/resources/documents/misc/57jpa8.htm> (accessed May 8, 2014).

⁵⁸ UN Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, Lethal Autonomous Robotics, pp. 21-22.

Claim #9: Fully autonomous weapons could have military and humanitarian benefits that would be sacrificed by a preemptive ban on such technology.

Response: The potential dangers posed by fully autonomous weapons would offset any possible benefits. Furthermore, delegating life-and-death decisions to machines would be deeply problematic from a moral standpoint.

Analysis: Critics argue that a preemptive ban on fully autonomous weapons would mean forgoing the potential advantages of such technology. They believe that the weapons could provide military benefits such as faster-than-human data processing and response time and enhanced protection of soldiers.⁵⁹ According to some, greater autonomy could also help prevent civilian harm through better target identification or more accurate estimations of the damage likely to result from a particular attack.⁶⁰

The drawbacks of such technology would outweigh the conceivable benefits, however. For example, the greater speed with which fully autonomous weapons might act, while arguably offering some military advantage, would raise the possibility that armed conflicts could rapidly spiral out of control. In arguing that fully autonomous weapons could become a necessity for states seeking to keep up with their adversaries, two critics write that “[f]uture combat may ... occur at such a high tempo that human operators will simply be unable to keep up. Indeed, advanced weapon systems may well create an environment too complex for humans to direct.”⁶¹ The frightening nature of such a scenario is compounded by the prospect of an arms race in which states increasingly feel the need to rely on fully autonomous weapons. The fact that some have gone so far as to contend that “autonomous weapons may become the norm on

⁵⁹ See, for example, Arkin, *Governing Lethal Behavior in Autonomous Robots*, p. 30 (Robots “can integrate more information from more sources far faster before responding with lethal force than a human possibly could in real-time.”); P.W. Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century* (New York: The Penguin Press, 2009), p. 418 (“[M]ost of the focus on military robotics is to use robots as a replacement for human losses.”).

⁶⁰ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 15 (“It may well be, for instance, that weapons systems with greater and greater levels of automation can—in some battlefield contexts, and perhaps more and more over time—reduce misidentification of military targets, better detect or calculate possible collateral damage, or allow for using smaller quanta of force compared to human decision-making.”). Schmitt and Thurnher, ““Out of the Loop,” *Harvard National Security Journal*, p. 3 (“Perhaps even more troubling is the prospect that banning autonomous weapon systems altogether based on speculation as to their future form could forfeit their potential use in a manner that would minimize harm to civilians and civilian objects when compared to non-autonomous weapon systems.”).

⁶¹ Schmitt and Thurnher, ““Out of the Loop,” *Harvard National Security Journal*, p. 238 (internal citation omitted).

the battlefield in a generation” makes it clear just how pressing the issue has become.⁶²

While perceived military advantage may be driving much of the push towards fully autonomous weapons, certain claims regarding such technology’s benefits are overstated. Critics correctly point out that countries have an interest in “fielding systems that enable them to deliver lethal force while minimizing the risk to their own forces,” and saving soldiers’ lives is a laudable goal.⁶³ Existing semi-autonomous systems, such as armed drones, however, provide similar force protection while leaving a human in control of the ultimate firing decision.

The use of fully autonomous weapons also raises the troubling specter that life-and-death decisions will be increasingly delegated to machines that, by their very nature, are without a moral compass. Thus, even if fully autonomous weapons might conceivably someday take actions with “potentially lethal consequences ... better than humans can”⁶⁴—a development Human Rights Watch and IHRC believe is unlikely—ceding control over lethal force to these weapons would be deeply problematic from a moral perspective. According to Christof Heyns, the UN special rapporteur on extrajudicial, summary or arbitrary executions, “[m]achines lack morality and mortality, and should as a result not have life and death powers over humans.”⁶⁵ Heyns adds that “[i]t is an underlying assumption of most legal, moral and other codes that when the decision to take life or to subject people to other grave consequences is at stake, the decision-making power should be exercised by humans.”⁶⁶ The fact that delegating life-and-death decisions to machines is so deeply disturbing across cultures should trump any arguable benefits touted by critics.

⁶² Ibid., p. 239 (describing the conclusions of certain US Department of Defense studies about future norms on the battlefield).

⁶³ Ibid., p. 232.

⁶⁴ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 3.

⁶⁵ UN Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, *Lethal Autonomous Robotics*, p. 17.

⁶⁶ Ibid., pp. 16-17.

Claim #10: Concerns about fully autonomous weapons would be best addressed through a regulatory approach rather than a ban.

Response: A binding, absolute ban on fully autonomous weapons would reduce the chance of misuse of the weapons, be easier to enforce, and enhance the stigma associated with violations.

Analysis: Certain critics object to a categorical ban on fully autonomous weapons because they prefer a regulatory framework that would permit the use of such technology in accordance with certain pre-defined limitations.⁶⁷ Such a framework might, for example, restrict the use of fully autonomous weapons to specific types of locations or purposes.⁶⁸ These critics suggest that such an approach would more precisely tailor restrictions to the evolving state of fully autonomous weapons technology and thus not be over-inclusive. Regulations could come in the form of a legally binding instrument or a set of gradually developed, informal standards.⁶⁹

An absolute, legally binding ban on fully autonomous weapons, however, would provide several distinct advantages over formal or informal constraints. It would maximize protection for civilians in conflict because it would be more comprehensive than regulation. A ban would also be more effective as it would prohibit the existence of the weapons and be easier to enforce. Moreover, a ban can have a powerful stigmatizing effect, creating a widely recognized new standard and influencing even those that do not join the treaty.

By contrast, once fully autonomous weapons came into being under a regulatory regime, they would be vulnerable to misuse. Even if regulations restricted use of fully autonomous weapons to certain locations or specific purposes, after the weapons entered national arsenals, countries that usually respect international humanitarian law could be tempted to use the weapons in inappropriate ways in the heat of battle or in dire circumstances. Furthermore, the existence of fully autonomous weapons would

⁶⁷ See Armin Krishnan, "Automating War: The Need for Regulation," *Contemporary Security Policy*, vol. 30, no. 1 (2009), p. 189 ("The best option of dealing with the possible implications of military robotics is probably not a general ban.... What is proposed in here as a solution is to allow defensive applications of [autonomous weapons], but to put considerable restrictions on offensive types and to ban certain types (self-evolving, self-replicating robots, microrobots) completely.").

⁶⁸ See, for example, *ibid.*, p. 188.

⁶⁹ Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems," *Jean Perkins Task Force on National Security and Law*, p. 22 (explaining that "[b]y 'international norms' here, we do not mean new binding legal rules only—whether treaty rules or customary international law—but instead the gradual fostering of widely-held expectations about legally or ethically appropriate conduct, whether formally binding or not").

leave open the door to their acquisition by repressive regimes or non-state armed groups that might disregard the restrictions or alter or override any programming designed to regulate a robot's behavior. They could use the weapons against their own people or civilians in other countries with horrific consequences.

Enforcement of regulations on fully autonomous weapons, as on all regulated weapons, could also be challenging and leave room for error, increasing the potential for harm to civilians. Instead of knowing that any use of fully autonomous weapons is unlawful, countries, international organizations, and nongovernmental organizations would have to monitor the use of the weapons and determine in every case whether use complied with the regulations. There would probably be debates about enforcement and the scope of the regulations—for example, what constituted a populated area, where use of certain weapons might be restricted.

The challenges of effectively controlling the use of fully autonomous weapons through binding regulations would be compounded if governments adopted a non-binding option. Those who support best practices advocate “let[ting] other, less formal processes take the lead to allow genuinely widely shared norms to coalesce in a very difficult area.”⁷⁰ To the extent that a “less formal” approach is a non-binding one, it is highly unlikely to constrain governments—including those already inclined to violate the law—in any meaningful way, especially under the pressures of armed conflict. It is similarly unrealistic to expect governments, as some critics hope, to resist their “impulses toward secrecy and reticence with respect to military technologies” and contribute to a normative dialogue about the appropriate use of fully autonomous weapons technology.⁷¹ If countries rely on transparency and wait until “norms coalesce” in an admittedly “very difficult area,”⁷² such weapons will likely be developed and deployed, at which point it would probably already be too late to control them.

⁷⁰ Ibid., p. 20.

⁷¹ Ibid., p. 25 (referring to the US tendencies toward secrecy).

⁷² Ibid., p. 20.

Claim #11: Efforts to ban fully autonomous weapons should be abandoned because an international prohibition is unrealistic.

Response: Past disarmament successes, growing support for a ban, and increasing international discussion of the issue suggest that a ban is a viable option for addressing fully autonomous weapons.

Analysis: Some critics argue that an absolute ban on the development, production, and use of fully autonomous weapons is “unrealistic.”⁷³ They have written that “part of our disagreements are about the practical difficulties that face international legal prohibitions of military technologies (we think such efforts are likely to fail).”⁷⁴ These critics fail to acknowledge the parallels with past successful disarmament efforts that had humanitarian benefits and the growing support for preserving meaningful human control over decisions to use lethal force.

Strong precedent exists for banning weapons that raise serious humanitarian concerns. The international community has previously adopted legally binding prohibitions on poison gas, biological weapons, chemical weapons, blinding lasers, and antipersonnel landmines. Most recently, 107 states adopted the 2008 Convention on Cluster Munitions, which comprehensively bans the use, production, transfer, and stockpiling of cluster munitions. Opponents of the landmine and cluster munitions instruments had frequently said that a ban treaty would never be possible, but the end results proved that their skepticism was misplaced.

Efforts to address the problems of fully autonomous weapons are following a similar path. April 2013 marked the launch of the Campaign to Stop Killer Robots, which calls for an absolute ban on the development, production, and use of fully autonomous weapons. The campaign resembles earlier civil society coalitions, including the Cluster Muniton Coalition and International Campaign to Ban Landmines.

Public support for a ban has bolstered the position of the campaign. In a 2013 public letter, more than 270 roboticists, artificial intelligence experts, and other scientists from 37 countries expressed doubts that adequate technological developments to ensure such weapons would comport with international humanitarian law would be

⁷³ Ibid., p. 3.

⁷⁴ Anderson and Waxman, “Human Rights Watch Report on Killer Robots, and Our Critique,” *Lawfare* blog, November 26, 2012, <http://www.lawfareblog.com/2012/11/human-rights-watch-report-on-killer-robots-and-our-critique/> (accessed April 15, 2014).

possible. They wrote, “[G]iven the limitations and unknown future risks of autonomous robot weapons technology, we call for a prohibition on their development and deployment. Decisions about the application of violent force must not be delegated to machines.”⁷⁵ Surveys have also revealed public resistance to the prospect of fully autonomous weapons. As mentioned under claim #7 above, a 2013 survey of 1,000 Americans found that two-thirds came out against the weapons, with almost three quarters of active duty military expressing opposition.⁷⁶

Interim measures to address the problem have also received attention. In May 2013, Christof Heyns, the UN special rapporteur on extrajudicial, summary or arbitrary executions, submitted to the Human Rights Council a report that was highly critical of fully autonomous weapons. He argued that the weapons would pose serious legal and ethical dangers and wrote that “[t]hey raise far reaching concerns about the protection of life during war and peace.”⁷⁷ Heyns called for national moratoria, which are temporary bans, on fully autonomous weapons and a high level panel to develop an international policy on the issue.⁷⁸ While not adopting the language of a moratorium, the US Department of Defense issued a de facto moratorium in a 2012 directive that prohibits for up to 10 years the use of fully autonomous systems to deliver lethal force.⁷⁹ Such moratoria can be a first step toward an absolute ban.

Finally, governments have taken up the debate about fully autonomous weapons in key international bodies. The Heyns report generated discussion in the Human Rights Council, where almost 20 nations expressed concern about the weapons and some endorsed the special rapporteur’s call for moratoria.⁸⁰ More importantly, the 117 states parties to the CCW agreed to take up the issue in informal discussions in May 2014. While this development does not indicate universal support for a ban, it does reflect a willingness by many major military powers to discuss concerns about the use of fully autonomous weapons in a leading disarmament forum. The CCW process has in the past produced a preemptive ban on blinding lasers and served as an incubator for bans on landmines and cluster munitions. There is certainly much work to be done to

⁷⁵ Noel Sharkey, “Computing Experts from 37 Countries Call for Ban on Killer Robots,” *International Committee for Robot Arms Control*, October 16, 2013, <http://icrac.net/2013/10/computing-experts-from-37-countries-call-for-ban-on-killer-robots/> (accessed February 16, 2014).

⁷⁶ Carpenter, “US Public Opinion on Autonomous Weapons.”

⁷⁷ UN Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, *Lethal Autonomous Robotics*, p. 1.

⁷⁸ *Ibid.*

⁷⁹ US Department of Defense, “Autonomy in Weapon Systems,” Directive 3000.09, November 21, 2012, pp. 3-4.

⁸⁰ Campaign to Stop Killer Robots, “Consensus Killer Robots Must be Addressed” May 28, 2013, <http://www.stopkillerrobots.org/2013/05/nations-to-debate-killer-robots-at-un/> (accessed April 15, 2014).

achieve a ban, but past precedents and recent developments suggest it cannot be dismissed as unrealistic.

Claim #12: A ban on the development of fully autonomous weapons would impede development of other valuable autonomous technology.

Response: A prohibition would not stifle such advances in autonomous technology because it would not cover non-weaponized, fully autonomous technology or semi-autonomous weapon systems.

Analysis: Some critics worry about the breadth of a ban on development. They express concern that it would represent a prohibition “even on the development of technologies or components of automation that could lead to fully autonomous lethal weapon systems.”⁸¹ These critics fear that the ban would therefore impede the exploration of beneficial autonomous technology, such as self-driving cars.

In fact, the ban would apply to development only of fully autonomous *weapons*, that is, machines that could select and fire on targets without human intervention. Research and development activities would be banned if they were directed at technology that can be used exclusively for fully autonomous weapons or that is explicitly intended for use in such weapons. A prohibition on the development of fully autonomous weapons would in no way impede development of fully autonomous robotics technology, which can have many positive, non-military applications.

The prohibition would also not encompass development of *semi*-autonomous weapons such as existing remote-controlled armed drones. Even if the prohibition is a narrow one, as a matter of principle, countries would not, and should not, be permitted to contract specifically for the development of fully autonomous weapon systems.

⁸¹ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 14.

Conclusion

Fully autonomous weapons raise a host of humanitarian concerns that cumulatively outweigh any potential benefits. They would face obstacles to complying with international humanitarian law, have the potential to proliferate and start an arms race, create an accountability gap, undermine non-legal checks on killing, and present the ethical problem of ceding life-and-death decisions to machines. To address these concerns, governments should adopt new international law to supplement existing international humanitarian law. In particular, they should agree to a preemptive ban, rather than regulation. An absolute prohibition on development, production, and use of fully autonomous weapon would be the most effective approach to eliminating the humanitarian threats posed by fully autonomous weapons.