The Need for New Law to Ban Fully Autonomous Weapons: Memorandum to Convention on Conventional Weapons Delegates

Human Rights Watch and
Harvard Law School's International Human Rights Clinic
November 2013

Introduction

International attention to the subject of fully autonomous weapons has grown rapidly over the past year. These weapons, also called "lethal autonomous robots" or "killer robots," would be able to identify and fire on targets without meaningful human intervention. Although they do not yet exist, they have generated widespread concern about their implications for the protection of civilians and combatants from unlawful attacks during armed conflict.

Several important developments have taken place since the November 2012 publication of *Losing Humanity: The Case against Killer Robots* by Human Rights Watch and Harvard Law School's International Human Rights Clinic (IHRC).[1]

- In November 2012, the United States became the first country to issue a public policy on these weapons, generally prohibiting their use for a period of five to 10 years despite certain loopholes.[2]
- April 2013 saw the launch of the Campaign to Stop Killer Robots, an international coalition of nongovernmental organizations, coordinated by Human Rights Watch, that calls for a preemptive ban on fully autonomous weapons.
- In May 2013, UN special rapporteur on extrajudicial killings, Christof Heyns, presented to the Human Rights Council his report recommending national

[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-0.

[2] US Department of Defense Directive, "Autonomy in Weapon Systems," no. 3000.09, November 21, 2012.

moratoria on these weapons until international discussions can be held.[3] The report prompted two dozen nations to speak about fully autonomous weapons for the first time at the United Nations.

In tandem with these high profile developments, there has been a rapidly expanding debate about fully autonomous weapons in diplomatic, legal, scientific, and military communities.

As of November 1, 2013, at least 30 nations had expressed concern about fully autonomous weapons, with many calling for urgent international discussions on the topic.[4]

This memorandum calls on states parties to the Convention on Conventional Weapons (CCW) to take up this challenge by adopting a mandate to discuss the issue in 2014 with an eye to negotiating a new protocol as quickly as possible.

This memorandum argues that existing international humanitarian law is insufficient to deal with fully autonomous weapons and provides four principle reasons why a supplementary legally binding instrument is needed:

- Fully autonomous weapons raise the kinds of concerns under the Martens Clause—which requires weapons to meet the "principles of humanity" and "dictates of public conscience"—that have justified the creation of past treaties.
- While existing international humanitarian law focuses on use, a new convention could also address development and production.
- Such a convention could help close the accountability gap associated with fully autonomous weapons.
- The precautionary principle, which applies perfectly to this situation, allows for action to prevent harm even though scientific uncertainty remains.

The memorandum concludes by spelling out why new law should take the form of a preemptive ban and laying out recommendations for global and domestic action.

---

[3] UN Human Rights Council, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns, A/HRC/23/47, April 9, 2013 [hereinafter Heyns Report].

[4] According to the Campaign to Stop Killer Robots, at least 30 nations have spoken publicly on fully autonomous weapons since the Heyns report was presented to the Human Rights Council on May 30, 2013: Algeria, Austria, Argentina, Belarus, Brazil, China, Costa Rica, Cuba, Ecuador, Egypt, France, Germany, Greece, India, Indonesia, Iran, Ireland, Japan, Mexico, Morocco, the Netherlands, New Zealand, Pakistan, Russia, Sierra Leone, South Africa, Sweden, Switzerland, the United Kingdom, and the United States. See: Campaign to Stop Killer Robots, "More States Speak out at UN," October 30, 2013, http://www.stopkillerrobots.org/2013/10/unga2013/ (accessed November 3, 2013).

Properties of Fully Autonomous Weapons

All robots possess a degree of autonomy, in other words, the ability to operate without human supervision. The exact level of autonomy, however, can vary greatly. Unmanned robotic weapons are often divided into three categories based on the amount of human involvement in their actions:

- Human-*in*-the-Loop Weapons: robots that can select targets and deliver force only with a human command;
- Human-*on*-the-Loop Weapons: robots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions; and
- Human-*out-of*-the-Loop Weapons: robots that are capable of selecting targets and delivering force without any human input or interaction.

For our purposes, the term "fully autonomous weapon" refers to both unmanned weapons that can target and deliver force without any human input, and those that can target and deliver force under the oversight of a human operator where such supervision is so limited the weapons are effectively "out of the loop."[5]

Defenders of fully autonomous weapons argue they have a range of potential benefits. They contend that these weapons would reduce the risks to soldiers and could increase accuracy of attacks and speed of response.[6] In addition, they note that pain, hunger, exhaustion, the instinct for self-defense, and emotions such as fear and anger would not bias fully autonomous weapons' determinations about when to use lethal force.

As this memorandum argues, however, the potential advantages of fully autonomous weapons would be offset by the lack of human control over the weapons. They would lack compassion, a check on the killing of other human beings. They also would face challenges in being able to comply with the complex and subjective rules of international humanitarian law.

These legal and policy concerns, elaborated on below, should be addressed in a new international treaty.

---

[5] Major Jeffrey Thurnher, a U.S. Army lawyer, notes the importance of a meaningful override. He writes that such robots "should be required to have some version of a human override" and that "[t]his oversight would not be effective if the human operator were merely a rubber stamp to approve an engagement." Jeffrey S. Thurnher, "No One at the Controls: Legal Implications of Fully Autonomous Targeting," *Joint Force Quarterly*, issue 67 (2012), p. 83.

[6] See, e.g., Michael Schmitt, "'Out of the Loop': Autonomous Weapons Systems and the Law of Armed Conflict," *Harvard National Security Journal*, vol. 4 (2013), pp. 232, 239.

Concerns under Core Principles of International Humanitarian Law

To be compliant with existing international humanitarian law, fully autonomous weapons, at a minimum, would need to be able to meet two bedrock principles: distinction and proportionality. These principles provide civilians essential protections in armed conflict. It is questionable whether fully autonomous weapons would be able to comply meaningfully with either one.[7]

The principle of distinction requires that a belligerent distinguish between combatants and civilians. Armed forces may lawfully attack individuals directly participating in hostilities, but must spare noncombatants.[8] Making these distinctions has become increasingly difficult because in many theaters of contemporary armed conflict, combatants often do not wear uniforms or insignia and seek deliberately to blend in with civilian populations. Combatants must be identified by their conduct rather than their appearance.

The ability to distinguish combatants and civilians rests not only on readily visible or audible signals, but also on judgment of an individual's intentions. While a machine would rely upon sensors and external stimuli to make such status determinations, a human soldier could consider numerous subtle and intuitive cues, including his or her perception of the individual's emotional state. There seems to be little prospect that in the near future, or perhaps ever, robots could be programmed to have the innately human qualities crucial to assessing an individual's intentions. This inability to determine intention could also undermine protection for soldiers, such as those wounded or surrendering, whom robots might not able to distinguish from active fighters.

A robot's lack of human judgment and intuition could similarly present an obstacle to compliance with the rule of proportionality, which prohibits attacks in which expected civilian harm outweighs anticipated military gain. Because proportionality relies heavily on situational and contextual factors, the lawful response to a situation could change considerably by slightly altering the facts. According to the US Air Force, "Proportionality in attack is an inherently subjective determination that will be resolved on a case-by-case basis."[9]

---

[7] While this paper focuses on international humanitarian law and situations of armed conflict, fully autonomous weapons also have the potential to raise concerns under human rights law.

[8] 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. 51(3); Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) , 1125 U.N.T.S. 609, entered into force December 7, 1978, art. 13(3).

[9] Air Force Judge Advocate General's Department, "Air Force Operations and the Law: A Guide for Air and Space Forces," first edition 2002, http://web.law.und.edu/Class/militarylaw/web_assets/pdf/AF%20Ops%20&%20Law.pdf (accessed October 28, 2013), p. 27.

It would be difficult to pre-program a robot to handle the infinite number of scenarios it might face. In addition, international humanitarian law depends on human judgment to make subjective decisions about the proportionality of attacks, and proportionality is ultimately "a question of common sense and good faith for military commanders."[10] Some experts doubt that a roboticist could equip a robot with metrics that would enable it to weigh the value of different military targets relative to risk of harm to different civilians and civilian objects in complex and evolving situations over time.[11]

In addition to raising questions about compliance with distinction and proportionality, fully autonomous weapons also challenge a foundational value of international law: human control over life-and-death decisions on the battlefield. According to the Heyns Report to the Human Rights Council, "It 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."[12]

Why a New Law?

Fully autonomous weapons represent a new category of weapons that could change the way wars are fought and pose serious risks to civilians. They, therefore, demand clarification and strengthening of existing international law. While international humanitarian law already sets limits on problematic weapons and their use, responsible countries have found it necessary to supplement that legal framework for several weapons that significantly threaten civilians, including antipersonnel mines, biological weapons, chemical weapons, and cluster munitions. Additionally, there is precedent for a preemptive ban on a class of weapons. In 1995, CCW states parties agreed to a ban on blinding lasers before the weapons had started to be deployed out

---

[10] International Committee of the Red Cross (ICRC), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Geneva: Martinus Nijhoff Publishers, 1987), http://www.icrc.org/ihl.nsf/COM/470-750073?OpenDocument (accessed October 28, 2013), pp. 679, 682.

[11] "[P]roportionality cannot be converted into an algorithmic formula necessary for autonomy because, at some point, a human has to be able to express it in common, measurable values." David Akerson, "The Illegality of Offensive Lethal Autonomy," in Dan Saxon, ed., *International Humanitarian Law and the Changing Technology of War* (Leiden: Martinus Nijhoff, 2013), p. 85. See also Noel Sharkey, "Killing Made Easy: From Joysticks to Politics," in Patrick Lin, Keith Abney, and George A. Bekey, eds., *Robot Ethics: The Ethical and Social Implications of Robotics* (Cambridge, MA: The MIT Press, 2012), p. 123. Roboticist Sharkey writes that determining proportionality "requires human judgment. No clear objective means are given in any of the laws of war for how to calculate what is proportionate…. What could the metric be for assigning value to killing an insurgent, relative to the value of noncombatants, particularly children…? The military says it is one of the most difficult questions that a commander has to make, but that acknowledgment does not answer the question of what metrics should be applied."

[12] Heyns Report, p. 16.

of concerns for the humanitarian harm the weapons would cause.[13] Fully autonomous weapons, even if likely to cause conventional types of injury, would have the potential to raise an especially high level of humanitarian concern, given that they would revolutionize warfare with an unprecedented shift away from human involvement. The following sections discuss several reasons to add to existing international law: the Martens Clause, the risks of an arms race and proliferation, an accountability gap, and the precautionary principle.

Martens Clause

In addition to presenting potential difficulties in meeting the core principles of proportionality and distinction, the use of fully autonomous weapons would raise concerns under the Martens Clause. A new instrument could help address these concerns.

The Martens Clause, which first appeared in the preamble to the Hague Conventions of 1899, is generally considered customary international law.[14] As articulated in Additional Protocol I to the Geneva Conventions, it states:

> In cases not expressly governed 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.[15]

The Martens Clause counters the assumption that anything not explicitly prohibited by relevant treaties is therefore permitted.[16] It establishes that in the face of continuing developments in technology and methods of warfare, the fundamental principle that laws of war should conform with custom, humanity, and the public conscience always holds.[17] Even though fully autonomous weapons are still in development, they should

---

[13] CCW Protocol IV on Blinding Laser Weapons, adopted October 13, 1995, entered into force July 30, 1998. See also ICRC, "Ban on Blinding Laser Weapons Now in Force," July 30, 1998, http://www.icrc.org/eng/resources/documents/misc/57jpa8.htm (accessed November 3, 2013).

[14] G.I.A.D Draper, "The Development of International Humanitarian Law," in UNESCO, *International Dimensions of Humanitarian Law* (Paris: UNESCO, 1988), pp. 72-73. See also International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, Dissenting Opinion of Judge Shahabuddeen, July 8, 1996, p. 183. The Martens Clause has since been incorporated, either specifically or by reference, into numerous treaties including Article 1(2) of the 1977 Additional Protocol I, and the preamble to the CCW. See also 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.

[15] Protocol I, art. 1(2).

[16] Jean Pictet, *Commentary on the Geneva Conventions* (Geneva: International Committee of the Red Cross, 1952), p. 32.

[17] Some scholars and judges have referred to the principles of the Martens Clause as a source of positive law, as opposed to merely considerations that are already part of the proportionality, distinction, and military necessity

be assessed according to the principles set forth by the Martens Clause. Where, as with fully autonomous weapons, there are risks that a weapon's use would contravene these basic considerations, international humanitarian law should be strengthened and clarified through a new binding international instrument.

Potential Concerns

It is doubtful that the use of fully autonomous weapons could fully comply with the principles of humanity. While there is no universal definition of this phrase, according to the International Committee of the Red Cross (ICRC), "the principle of humanity is perfectly natural: it is compassion, mutual aid, a reaching out to others to help and protect them."[18] The ICRC adds that the principle "embodies one especially important idea: to protect."[19] The use of fully autonomous weapons would likely be inconsistent with the principle of humanity's values of compassion and protection.

While, as mentioned above, fully autonomous weapons would not share the emotional weaknesses of human soldiers, they would at the same time be bereft of other emotions, most notably compassion. Compassion can deter combatants from killing other human beings even in conflicts where there is little regard for international humanitarian law and commanders order troops to target civilians. In addition, the use of fully autonomous weapons could undermine the ability to protect civilians. For the reasons already discussed, these weapons would face challenges in complying with the foundational rules of distinction and proportionality, which are designed to protect civilians in armed conflict.

The use of fully autonomous weapons could also run counter to the dictates of public conscience. The definition of the public conscience is debated, but public opinion can play an important role in both revealing and shaping it.[20] Although few countries have promulgated official positions on fully autonomous weapons, for many people the

---

calculus. Judge Shahabudden stated that the Martens Clause is not a mere reflection of the existence of customary law outside of specific treaties, but rather "provide[s] authority for treating the principles of humanity and the dictates of public conscience as principles of international law." International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, p. 184.

[18] ICRC, "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.

[19] Ibid. Others have interpreted the principle as "prohibiting means and methods of war which are not necessary for the attainment of a definite military advantage." Rupert Ticehurst, "The Martens Clause and the Laws of Armed Conflict," *International Review of the Red Cross*, vol. 317 (1997), http://www.icrc.org/eng/resources/documents/misc/57jnhy.htm (accessed October 20, 2013). A review of cases and other sources that reference the principles of humanity suggests, however, that the ICRC's definition is more in line with the underlying concerns that the principles of humanity address. See, for example,. *The Corfu Channel Case (Merits)*, ICJ Reports, 1949 (elementary considerations of humanity); *United States of America v. Alfried Krupp, et al.*, U.S. Military Tribunal Nuremberg, 1948, pp. 14-15; Vincent Bernard, ed., "New Technologies and Warfare," *International Review of the Red Cross*, vol. 94 (2012).

[20] See Theodor Meron, "The Martens Clause, Principles of Humanity, and Dictates of Public Conscience," *American Journal of International Law*, vol. 95 (2000), p. 83.

prospect of these weapons is disturbing. In discussions with government and military officials, scientists, and the general public, for example, Human Rights Watch has encountered tremendous discomfort with the idea of allowing military robots to determine on their own if and when to use lethal force against a human being.

A June 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 development. Interestingly, active duty military personnel were among the strongest objectors—73 percent expressed opposition to fully autonomous weapons.[21] Special rapporteur Christof Heyns articulated the public concern about crossing an ethical line in his report to the Human Rights Council: "Machines lack morality and mortality, and should as a result not have life and death powers over humans."[22]

Prior Use of 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 widespread condemnation of poison gas was highlighted in the text of the Geneva Gas Protocol of 1925. The protocol's preamble states that gas "has been justly condemned by the general opinion of the civilized world."[23] The "general opinion of the civilized world" reflected the dictates of public conscience.[24]

The Martens Clause also heavily influenced the discussions and debates preceding the development of CCW Protocol IV on Blinding Lasers,[25] which bans the transfer and use of laser weapons whose sole or partial purpose is to cause permanent blindness.[26] Not

---

[21] Charli Carpenter, "US Public Opinion on Autonomous Weapons," June 19, 2006, http://www.whiteoliphaunt.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf (accessed June 21, 2013). 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.whiteoliphaunt.com/duckofminerva/2013/06/how-do-americans-feel-about-fully-autonomous-weapons.html (accessed June 21, 2013). 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/.

[22] Heyns Report, p. 17.

[23] Protocol for the Prohibition of the Use and Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, adopted June, 17, 1925, entered into force February 8, 1928.

[24] Human Rights Watch noted that "the use of poison gas in World War I led to the conclusion of the 1925 Gas Protocol, which draws on the Martens clause." Human Rights Watch, *Blinding Laser Weapons*.

[25] David Akerson, "The Illegality of Offensive Lethal Autonomy," in Saxon, ed., *International Humanitarian Law and the Changing Technology of War*, pp. 92-93.

[26] CCW Protocol IV on Blinding Lasers, adopted Oct. 13, 1995, entered into force July 30, 1998, art. 1.

only was the Martens Clause invoked by civil society in its reports on the matter,[27] but experts participating in a series of ICRC meetings on the subject also turned to considerations of humanity and public conscience in their analysis. They largely agreed that "[blinding lasers] would run counter to the requirements of established custom, humanity, and public conscience."[28] The use of the Martens Clause was significant because it was unclear to what extent blinding lasers were indiscriminate weapons or caused superfluous injury.[29] 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.[30] The Blinding Lasers Protocol set an international precedent for banning weapons based not only on considerations distinction, proportionality, and unnecessary suffering, but also widespread revulsion to their use.[31]

While fully autonomous weapons would encompass a broader range of specific types of weapons, there are parallels between the discussion about blinding lasers and the current debate on fully autonomous weapons. Similar concerns under the principles of humanity and the dictates of public conscience argue for an absolute and preemptive ban on fully autonomous weapons akin to that on blinding lasers.

The Risks of an Arms Race and Proliferation

Because international humanitarian law generally addresses only the *use* of weapons, supplemental law is needed to minimize the risks associated with the development and production of fully autonomous weapons, namely an arms race and proliferation. As countries invest more time and resources into developing fully autonomous weapons, it will be increasingly difficult to persuade them to give them up. In fact, as development of these weapons gathers pace, countries could have extra incentive to continue development. If one nation acquires these weapons, others may feel they have to follow suit to avoid falling behind in a robotic arms race. Such a race may

---

[27] See, for example, Human Rights Watch, *Blinding Laser Weapons*.

[28] 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." ICRC, *Blinding Weapons: Reports of the Meetings of Experts Convened by the International Committee of the Red Cross on Battlefield Laser Weapons, 1989-1991* (Geneva: ICRC, 1993), pp. 344-46. Others doubted their ability to field such weapons, notwithstanding possible military utility, because of public opinion. *Ibid*, p. 341.

[29] Akerson, "The Illegality of Offensive Lethal Autonomy," in Saxon, ed., *International Humanitarian Law and the Changing Technology of War*, pp. 92-93.

[30] 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." ICRC, *Blinding Weapons*, p. 345.

[31] Akerson, "The Illegality of Offensive Lethal Autonomy," in Saxon, ed., *International Humanitarian Law and the Changing Technology of War*, p. 96.

already be in its early stages. US Department of Defense studies "have recommended 'aggressively' incorporating autonomy into future systems," and they suggested "autonomous weapons may become the norm on the battlefield in a generation."[32] Some of the other countries pursuing ever-greater autonomy for weapons include China, Israel, Russia, South Korea, and the United Kingdom. A new instrument that banned development and production of fully autonomous weapons could prevent the escalation of an arms race.

If development of fully autonomous weapons is left unchecked, countries might argue that the principle of military necessity allows them to use these weapons to counter an adversary's comparable weapons. One scholar cautions, "Technology can largely affect the calculation of military necessity…. It might be necessary to restrict, or maybe even prohibit [autonomous weapons] from the beginning in order to prevent a dynamics [sic] that will lead to the complete automation of war that is justified by the principle of necessity."[33]

Finally, even if fully autonomous weapons were initially developed by governments that generally comply with international humanitarian law, the weapons' existence would open the door to proliferation to repressive regimes or non-state armed groups with little regard for the law. Fully autonomous weapons could be perfect tools of repression for autocrats seeking to strengthen or retain power. Even the most hardened troops can eventually turn on their leader if ordered to fire on their own people. An abusive leader who resorted to fully autonomous weapons would be free of the fear that armed forces would resist being deployed against certain targets.

To address concerns about proliferation, many stand-alone disarmament treaties have included prohibitions on development and production as well as on use. The Biological Weapons Convention, Chemical Weapons Convention, Mine Ban Treaty, and Convention on Cluster Munitions all include such bans. An instrument on fully autonomous weapons should follow that precedent.

Accountability

There are serious concerns not only about fully autonomous weapons' inability to comply with existing international humanitarian law but also about the lack of accountability when they fail to do so. A new legally binding instrument that explicitly bans the weapons would eliminate that accountability gap.

---

[32] Schmitt, "Out of the Loop," *Harvard National Security Journal,* pp. 238-239, citing U.S. Joint Forces Command, "Unmanned Effects (UFX): Taking the Human Out of the Loop," Rapid Assessment Process (RAP) Report #03-10 (2003), http://edocs.nps.edu/dodpubs/org/JFC/RAPno.03-10.pdf.

[33] Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Surrey, UK: Ashgate Publishing Limited, 2009), p. 91-92 (quoted in Human Rights Watch and IHRC, *Losing Humanity*, p. 35).

Accountability for violations of international humanitarian law is important for two reasons. First, accountability deters individuals (and leaders with command responsibility) from committing war crimes or otherwise failing to adhere to the laws of war. Second, accountability for unlawful acts dignifies victims by giving them recognition that they were wronged and satisfaction that someone was punished for inflicting the harm they experienced.

Holding a human responsible for the actions of a robot that is acting autonomously could prove difficult. For example, the doctrine of "command responsibility" holds commanders legally responsible for their subordinates' violations of international humanitarian law in limited circumstances: it applies if the commander knew or should have known that the individual planned to commit a crime yet he or she failed to take action to prevent it or did not punish the perpetrator after the fact.[34] This doctrine is ill suited for fully autonomous weapons. It would appropriately create culpability for a commander who recklessly deployed fully autonomous weapons in a way he or she knew would endanger civilians; however, in many cases it would leave an accountability gap. A commander would be unlikely to be held legally responsible if he or she were unable to identify a danger prior to deployment because he or she had not programmed the robot and could not necessarily predict its behavior. A commander could not punish a robot because the robot itself could not be meaningfully held accountable either for the purposes of deterrence or as a means of providing justice for victims.

Holding a programmer accountable for deficiencies in a robot's judgment or responses could also be difficult; absent proof of intentionality, criminal law would be unlikely to apply, and many government weapons-developers would be protected from even civil liability except in the most egregious circumstances.[35] Treating a robot-produced civilian massacre as though it were a product liability situation rather than a subject for

---

[34] Protocol I, arts. 86(2), 87.

[35] The US Supreme Court, for example, has established the "government contractor defense" that generally excuses government contractors from civil liability for the design of weapons. *Boyle v. United Techs. Corp.* states:

> Liability for design defects in military equipment cannot be imposed, pursuant to state law, when (1) the United States approved reasonably precise specifications; (2) the equipment conformed to those specifications; and (3) the supplier warned the United States about the dangers in the use of the equipment that were known to the supplier but not to the United States.
>
> *Boyle v. United Techs. Corp.*, 487 U.S. 500, 512 (1988).

US courts have also applied the "combatant activity exclusion" to excuse weapons designers and manufacturers from civil liability arising out of incidents occurring during the course of war. *Koohi v. United States*, 976 F.2d 1329 (9th Cir. 1992) (holding that a design defect claim against Aegis manufacturers arising from the shooting down of a civilian aircraft by an United States warship was precluded on the grounds that "the imposition of such liability on the manufacturers of the Aegis would create a duty of care where the combatant activities exception is intended to ensure that none exists").

criminal prosecution could reduce incentives to exercising the highest level of care with respect to international humanitarian law compliance. Programmers and manufacturers might find that the potential benefits or profits of fully autonomous weapons outweigh the risk of civil liability for a massacre.

This tangled prospect for accountability under international humanitarian law requires a new international instrument prohibiting fully autonomous weapons. A ban treaty would aim to end use and thus eliminate situations in which it was impossible to achieve effective deterrence or retributive justice for breaches of the laws of war involving fully autonomous weapons. The treaty would be explicit that it would "never under any circumstances" be lawful to use a fully autonomous weapon and that anyone violating that rule would be held responsible for the weapon's actions.[36]

Precautionary Principle

While fully autonomous weapons raise many legal and ethical concerns, some defenders argue there is no proof that a technological fix could not solve the problem. There is an equal lack of proof, however, that a technological fix would work. Given the concerns these weapons raise, the scientific uncertainty that exists, and the potential benefits of a new legally binding instrument, the precautionary principle is directly applicable. This principle suggests that the international community need not wait for scientific certainty, but could and should take action now.

The precautionary principle, a general principle of international law, applies when a party is considering an act or policy that has the potential to cause public harm.[37] If it is uncertain whether or not the act will in fact be harmful, the party committing the act bears the burden of proof to demonstrate that the act will not be harmful.[38] Perhaps the most generally accepted formulation of the precautionary principle appears in the Rio Declaration, a product of the 1992 United Nations Conference on Environment and Development.[39] The Declaration's seminal Principle 15 states: "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. 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

---

[36] The Mine Ban Treaty and the Convention on Cluster Munitions both use the phrase "never under any circumstances" in their prohibitions.#

[37] Timothy O'Riordan and James Cameron, *Interpreting the Precautionary Principle* (London: Cameron May, 1994), p. 18.

[38] Timothy O'Riordan and James Cameron, *Reinterpreting the Precautionary Principle* (London: Cameron May, 2001), p. 20.

[39] Philip M. Kannan, "The Precautionary Principle: More than a Cameo Appearance in United States Environmental Law?" *William & Mary Environmental Law & Policy Review*, vol. 31 (2007), p. 420.The Conference, held in Rio de Janeiro in 1992, was intended to address 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 November 3, 2013).

measures to prevent environmental degradation."[40] Principle 15 thus clarifies that scientific uncertainty cannot be used as an excuse to avoid taking action.

Rio Principle 15 refers expressly to "environmental degradation," but its language can be adapted to other kinds of harm. The precautionary principle is directly applicable to fully autonomous weapons because there are (1) "threats of serious or irreversible damage," (2) "lack of full scientific certainty," and (3) "cost-effective measures to prevent [harm]."

According to the Rio Declaration, the precautionary principle is triggered by the threat of serious or irreversible damage. The development, production, and use of fully autonomous weapons certainly present a threat to civilians that would be both serious and irreversible, as the technology, once developed and employed, is unlikely to be effectively eliminated or easily suppressed.

Under the precautionary principle, it is not necessary to resolve scientific uncertainty in order for preventive measures to be warranted. At this time, there is no absolute proof as to whether one or more technological improvements could eliminate the threat posed by fully autonomous weapons. Some defenders have argued that a more certain scientific picture should be allowed to emerge before measures against the weapons are taken. Today's scientific uncertainty combined with the potential threat to the civilian population, however, suffices to open the door to immediate preventive measures.

The Rio formulation of the precautionary principle requires that preventive measures be "cost-effective," in other words, reasonable in cost relative to the anticipated gain. The costs of creating new law to govern fully autonomous weapons would need to be measured against the benefits it might ultimately provide. While treaty negotiations and implementation would carry costs, these expenses are small compared to the significant harm they might prevent. Moreover, a ban on such weapons would likely be more cost-effective to monitor and implement than efforts to regulate the development, proliferation and deployment of such weapons; as discussed below, bans are easier and arguably cheaper to enforce, and countries would not spend money on acquisition and deployment of the weapons. The cost of a ban on fully autonomous weapons should thus not be a barrier to its adoption.

While the influential Rio version of the precautionary principle does not *compel* preventive measures in the face of scientific uncertainty, other formulations of the principle call more strongly for preventive measures against harm such as that associated with fully autonomous weapons. For example, the Wingspread Consensus

[40] Rio Declaration on Environment and Development, adopted June 14, 1992, UN Doc. A/CONF.151/26 (vol. I) / 31 ILM 874 (1992), Principle 15.

Statement on the Precautionary Principle, a document promulgated by a group of treaty negotiators, activists, scholars, and scientists, provides: "When an activity raises threats of harm to human health or the environment, precautionary measures *should be taken* even if some cause and effect relationships are not fully established scientifically."[41] Even under the more modest Rio approach, however, the grave concerns outlined in this paper provide ample reason to undertake preventive measures in the form of an absolute ban.

Why An Absolute and Preemptive Ban

While international humanitarian law provides important limits on problematic weapons and their use, there are compelling reasons, outlined above, for a new international instrument on fully autonomous weapons. That treaty should establish an absolute ban on the development, production, and use of these weapons.

The ban would apply to any fully autonomous weapon, a weapon that could select and fire on targets without meaningful human intervention. Treaty drafters would have to consider what constitutes human intervention as they crafted a definition. A ban would not, however, entail a prohibition on all development of autonomous robotic technology. Research and development activities should be banned if they are directed at technology that can only be used for fully autonomous weapons or that is explicitly intended for use in such weapons. The prohibition would also not encompass development of semi-autonomous weapons, like today's remote-controlled drones, because those weapons systems that are not fully autonomous raise different kinds of concerns. Although the ban on development might be narrow, as a matter of principle, countries should not be permitted to undertake or contract research specifically for the development of fully autonomous weapons systems.

An absolute international prohibition of fully autonomous weapons would maximize protection from these weapons for civilians in conflict. It would be more comprehensive than regulating the use of these weapons, eliminate the need for case-by-case determinations, and make it easier to standardize rules across countries. Even if regulations restricted use of fully autonomous weapons to certain locations or to specific purposes, after the weapons entered national arsenals, countries might be tempted to use them in inappropriate ways in the heat of battle or in dire circumstances. Moreover, a ban can have a powerful stigmatizing effect, raising the political stakes for countries that use prohibited weapons.

[41] Wingspread Consensus Statement on the Precautionary Principle, January 26, 1998, http://www.sehn.org/wing.html (accessed November 3, 2013) (emphasis added).

A ban would also be more effective as it would be clearer and simpler to enforce. Enforcement of regulations, by contrast, can be challenging and leave room for error, increasing the potential for harm caused by fully autonomous weapons. Instead of clearly understanding that any use of fully autonomous weapons is unlawful, the international community would have to monitor the weapons' use and determine whether the use of a particular weapon complied with the regulations. There would probably be debates surrounding the scope of the regulations and enforcement.

The ban should be preemptive. It is difficult to stop technology once large-scale investments have been made. Even if these weapons were deployed and the concerns about their potential harm to civilians proved justified, it would be difficult to put the genie back in the bottle. 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.

Additionally, a preemptive ban would avoid the risks that the process of technological development might pose to human lives. Defenders of fully autonomous weapons argue countries should be allowed to pursue development of artificial intelligence for these weapons because eventually the technical concerns of opponents might be met. The process of ongoing development, however, is incremental learning from mistakes. If ongoing development is permitted, there will be a danger that fully autonomous weapons will be deployed in complex situations before the artificial intelligence to handle such situations sufficiently is achieved. The development of these weapons would be a process of trial and error. It is likely that only after facing unanticipated situations that were not previously programmed could the technology be developed to address those issues. During that period, the weapon would be particularly prone to mishandling a situation and causing civilian harm.

Some defenders of fully autonomous weapons contend that these weapons would advance incrementally but that their existence would be inevitable and therefore require regulation.[42] Indeed some states might be motivated to continue development even if a ban on the fully autonomous weapons was adopted. Nevertheless, a preemptive ban is the approach that would maximize humanitarian protection. The stigma generated by a ban would likely make it difficult for even rogue actors to deploy them without facing a problem of global illegitimacy. For example, Syria had incentive to disarm after international criticism of its use of chemical weapons.

[42] Kenneth Anderson and Matthew Waxman, "Law and Ethics for Robot Soldiers," *Policy Review* (Hoover Institution, Stanford University), no. 176, December 1, 2012, http://www.hoover.org/publications/policy-review/article/135336 (accessed November 3, 2013).

Recommendations

To minimize the threats fully autonomous weapons would likely pose to civilians, Human Rights Watch and IHRC recommend that:

- *All countries ban fully autonomous weapons through an international legally binding instrument.* Such a ban should be preemptive and cover development, production, and use of the weapons.

- *CCW states parties adopt a mandate to discuss fully autonomous weapons over the course of 2014.* We call for the adoption of a CCW mandate to hold constructive discussions on fully autonomous weapons with an eye to future action, in particular negotiation of a new instrument prohibiting fully autonomous weapons. The mandate should be broad enough to address the range of issues surrounding development, production, and use of fully autonomous weapons and reflect a sense of urgency and purpose.

- *All countries should adopt national laws and policies to prohibit the development, production, and use of fully autonomous weapons.* First steps toward an international, preemptive ban could be national laws prohibiting fully autonomous weapons or, as the UN special rapporteur on extrajudicial killings called for in his May 2013 report, national moratoria that lead to national or international bans. These actions would ensure that problematic weapons do not come into being and are not deployed while countries engage in a negotiation process for an international treaty. Moratoria could also be adopted by nations that would not join a treaty immediately, but wanted to place prohibitions on fully autonomous weapons at the domestic level in the interim.