

MAKING THE CASE

The Dangers of Killer Robots and the Need for a Preemptive Ban

HUMAN
RIGHTS
WATCH

IHRC

INTERNATIONAL HUMAN RIGHTS CLINIC
HUMAN RIGHTS PROGRAM AT HARVARD LAW SCHOOL



IHRC



Making the Case

The Dangers of Killer Robots and the Need for a Preemptive Ban

Copyright © 2016 Human Rights Watch

All rights reserved.

Printed in the United States of America

ISBN: 978-1-6231-34310

Cover design by Rafael Jiminez

Cover cartoon by Russell Christian

Human Rights Watch defends the rights of people worldwide. We scrupulously investigate abuses, expose the facts widely, and pressure those with power to respect rights and secure justice. Human Rights Watch is an independent, international organization that works as part of a vibrant movement to uphold human dignity and advance the cause of human rights for all.

Human Rights Watch is an international organization with staff in more than 40 countries, and offices in Amsterdam, Beirut, Berlin, Brussels, Chicago, Geneva, Goma, Johannesburg, London, Los Angeles, Moscow, Nairobi, New York, Paris, San Francisco, Sydney, Tokyo, Toronto, Tunis, Washington DC, and Zurich.

For more information, please visit our website: <http://www.hrw.org>

The International Human Rights Clinic (IHRC) at Harvard Law School seeks to protect and promote human rights and international humanitarian law through documentation; legal, factual, and strategic analysis; litigation before national, regional, and international bodies; treaty negotiations; and policy and advocacy initiatives. IHRC also critically examines the human rights movement and engages in innovative clinical education to develop advanced practice techniques and approaches to human rights advocacy. IHRC collaborates with leading international and local human rights organizations and bridges theory with practice at the law school while also advancing the interests of clients and affected communities around the world.

For more information, please visit IHRC's website: <http://hrp.law.harvard.edu/clinic/>



Making the Case

The Dangers of Killer Robots and the Need for a Preemptive Ban

Summary 1

Recommendations 3

I. The Dangers of Fully Autonomous Weapons..... 4

 Legal Dangers4

 Non-Legal Dangers 21

II. Arguments for a Preemptive Prohibition on Fully Autonomous Weapons 31

 Advantages of a Ban 31

 Timeliness and Feasibility of a Ban 40

Acknowledgments 49

Summary

The debate about fully autonomous weapons has continued to intensify since the issue reached the international stage four years ago.¹ Lawyers, ethicists, military personnel, human rights advocates, scientists, and diplomats have argued, in a range of venues, about the legality and desirability of weapons that would select and engage targets without meaningful human control over individual attacks. Divergent views remain as military technology moves toward ever greater autonomy, but there are mounting expressions of concern about how these weapons could revolutionize warfare as we know it. This report seeks to inform and advance this debate by further elaborating on the dangers of fully autonomous weapons and making the case for a preemptive ban.

In December 2016, states parties to the Convention on Conventional Weapons (CCW) will convene in Geneva for the treaty's Fifth Review Conference and decide on future measures to address "lethal autonomous weapons systems" (LAWS), their term for these weapons. Spurred to act by the efforts of the Campaign to Stop Killer Robots, CCW states have held three informal meetings of experts on LAWS since 2014. At the Review Conference, states parties should agree to establish a Group of Governmental Experts. The formation of this formal body would compel states to move beyond talk and create the expectation of an outcome. That outcome should be a legally binding prohibition on fully autonomous weapons.

¹ Human Rights Watch and the Harvard Law School International Human Rights Clinic helped spark discussions about fully autonomous weapons with their report *Losing Humanity: The Case against Killer Robots*, released in 2012. Since then they have produced a series of reports and papers on the topic. See, for example, Human Rights Watch and Harvard Law School International Human Rights Clinic (IHRC), *Losing Humanity: The Case against Killer Robots*, November 2012, <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>; "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>; "Fully Autonomous Weapons: Questions and Answers," October 2013, <http://www.hrw.org/news/2013/10/21/qa-fully-autonomous-weapons>; "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>; *Shaking the Foundations: The Human Rights Implications of Killer Robots*, May 2014, https://www.hrw.org/sites/default/files/reports/armso514_ForUpload_o.pdf; *Mind the Gap: The Lack of Accountability for Killer Robots*, April 2015, https://www.hrw.org/sites/default/files/reports/armso415_ForUpload_o.pdf; "Precedent for Preemption: The Ban on Blinding Lasers as a Model for a Killer Robots Prohibition," November 2015, https://www.hrw.org/sites/default/files/supporting_resources/robots_and_lasers_final.pdf; "Killer Robots and the Concept of Meaningful Human Control," April 2016, <https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control>.

To build support for a ban, this report responds to critics who have defended the developing technology and challenged the call for a preemptive prohibition. The report identifies 16 of the critics' key contentions and provides a detailed rebuttal of each. It draws on extensive research into the arguments on all sides. In particular, it examines academic publications, diplomatic statements, public surveys, UN reports, and international law.

The report updates a May 2014 paper, entitled "Advancing the Debate on Killer Robots," and expands it to address new issues that have surfaced over the past two years.² In the process, the report illuminates the major threats posed by fully autonomous weapons and explains the advantages and feasibility of a ban.

The first chapter of this report elaborates on the legal and non-legal dangers posed by fully autonomous weapons. The weapons would face significant obstacles to complying with international humanitarian and human rights law and would create a gap in accountability. In addition, the prospect of weapons that could make life-and-death decisions generates moral outrage, and even the expected military advantages of the weapons could create unjustifiable risks.

The second chapter makes the case for a preemptive prohibition on the development, production, and use of fully autonomous weapons. Of the many alternatives proposed, only an absolute ban could effectively address all the concerns laid out in the first chapter. The ban should be adopted as soon as possible, before this revolutionary and dangerous technology enters military arsenals. Precedent from past disarmament negotiations and instruments shows that the prohibition is achievable and would be effective.

² Human Rights Watch and IHRC, "Advancing the Debate on Killer Robots: 12 Key Arguments for a Preemptive Ban on Fully Autonomous Weapons," May 2014, <https://www.hrw.org/news/2014/05/13/advancing-debate-killer-robots>.

Recommendations

In light of the dangers posed by fully autonomous weapons and the inability to address these dangers other than with a ban, Human Rights Watch and the International Human Rights Clinic (IHRC) at Harvard Law School call on states to:

- Adopt an international, legally binding instrument that prohibits the development, production, and use of fully autonomous weapons;
- Adopt national laws or policies that establish prohibitions on the development, production, and use of fully autonomous weapons; and
- Pursue formal discussions under the auspices of the CCW, beginning with the formation of a Group of Governmental Experts, to discuss the parameters of a possible protocol with the ultimate aim of adopting a ban.

I. The Dangers of Fully Autonomous Weapons

Fully autonomous weapons raise a host of concerns. It would be difficult for them to comply with international law, and their ability to act autonomously would interfere with legal accountability. The weapons would also cross a moral threshold, and their humanitarian and security risks would outweigh possible military benefits. Critics who dismiss these concerns depend on speculative arguments about the future of technology and the false presumption that technological developments can address all of the dangers posed by the weapons.

Legal Dangers

Contention #1: Fully autonomous weapons could eventually comply with international humanitarian law, notably the core principles of distinction and proportionality.

Rebuttal: The difficulty of programming human traits such as reason and judgment into machines means that fully autonomous weapons would likely be unable to comply reliably with international humanitarian law.

Analysis: Some critics contend that fully autonomous weapons could comply with the core principles of distinction and proportionality, at some point in the future. They argue that advocates of a ban often “fail to take account of likely developments in autonomous weapon systems technology.”³ According to the 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 continue.⁴ Thus, while recognizing the existence of “outstanding issues” and “daunting problems,”⁵ critics are content with the belief that solutions are “theoretically achievable.”⁶ Proceeding on an assumption that

³ 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. 234.

⁴ 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 November 20, 2016), p. 11.

⁵ Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Raton, FL: CRC Press, 2009), pp. 126, 211.

⁶ Schmitt, “Autonomous Weapon Systems,” *Harvard National Security Journal Features*, p. 17 (discussing in particular whether autonomous weapons could be programmed adequately to “compute doubt”).

such weapons could one day conform to the international humanitarian law requirements of distinction and proportionality, however, is unwise.

Difficulties with 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.⁷ Although progress is likely in the development of sensory and processing capabilities, distinguishing an active combatant from a civilian or an injured or surrendering soldier requires more than such capabilities. It also depends on the qualitative ability to gauge human intention, which involves interpreting the meaning of subtle clues, such as tone of voice, facial expressions, or body language, in a specific context. Humans possess the unique capacity to identify with other human beings and are thus equipped to understand the nuances of unforeseen behavior in ways that machines, which must be programmed in advance, simply cannot. Replicating human judgment in determinations of distinction—particularly on contemporary battlefields where combatants often seek to conceal their identities—is a difficult problem, and it is not credible to assume a solution will be found.

Obstacles to Determining Proportionality

The obstacles to fully autonomous weapons complying with the principle of distinction would be compounded for proportionality, which requires the delicate balancing of two factors: expected civilian harm and anticipated military advantage. Determinations of proportionality take place not only in developing an overall battle plan, but also during actual military operations, when decisions must be made about the course or cessation of any particular attack. One critic concludes 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 While acknowledging that “it is unlikely in the near future that ... ‘machines’ will be programmable to perform robust assessments of a strike’s likely

⁷ The rule of distinction is required under both Additional Protocol I to the Geneva Conventions and under customary international law. See 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. 48; International Committee of the Red Cross (ICRC), *Customary International Humanitarian Law Database*, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule1 (accessed November 20, 2016), Rule 1.

⁸ Schmitt, “Autonomous Weapon Systems,” *Harvard National Security Journal Features*, p. 20.

military advantage,” he contends that “military advantage algorithms could *in theory* be programmed into autonomous weapon systems.”⁹

There are a number of reasons 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. When assessing proportionality, 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 attack’s impact on all entities and objects surrounding the target.

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 Military advantage must be determined on a “case-by-case” basis, and a programmer could not account in advance for the infinite number of unforeseeable 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, it 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 evaluating 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.

While some critics focus on the human commander’s action ahead of the strike,¹³ the proportionality of any particular attack depends on conditions at the time of the attack,

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*, International Criminal Tribunal for the Former Yugoslavia, http://www.difesa.it/SMD_/CASD/IM/ISSMI/Corsi/Corso_Consigliere_Giuridico/Documents/72470_final_report.pdf (accessed November 20, 2016), para. 50.

¹³ 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 sometimes have to be made

and not at the moment of design or deployment of a weapon. A commander weighing proportionality at the deployment stage would have to rely on the programmer's and manufacturer's predictions of how a fully autonomous weapon would perform in a future attack. No matter how much care was taken, a programmer or manufacturer would be unlikely accurately to anticipate a machine's reaction to shifting and unforeseeable conditions in every scenario. The decision to deploy a fully autonomous weapon is not equivalent to the decision to attack, and at the moment of making a determination to attack, such a weapon would not only be out of the control of a human being exercising his or her own judgment, but also unable to exercise genuine human judgment itself (see Contention #12).

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 note 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 (see Contention #6). Yet this is precisely what the reasonable military commander standard requires. Moreover, reasonableness eludes "objective definition" and depends on the situation.¹⁶

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 March 2013, <http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1679?prd=EPIL#law-9780199231690-e1679-div1-1> (accessed November 20, 2016), 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*, 2013, http://media.hoover.org/sites/default/files/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf (accessed November 20, 2016), p. 23.

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

Proportionality analyses allow for a “fairly broad margin of judgment,”¹⁷ but 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. As Christof Heyns, then UN special rapporteur on extrajudicial, summary or arbitrary executions, explained in his 2013 report, 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

While the capabilities of future technology are uncertain, it seems highly unlikely that it could ever replicate the full range of inherently human characteristics necessary to comply with the rules of distinction and proportionality. Adherence to international humanitarian law requires the qualitative application of judgment to what one scientist describes as an “almost indefinite combination of contingencies.”²⁰ Some experts “question whether artificial intelligence, which always seems just a few years away, will ever work well enough.”²¹

¹⁷ ICRC, *Commentary of 1987 on 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, <https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=D80D14D84BF36B92C12563CD00434FBD> (accessed November 20, 2016), art. 57, para. 2210.

¹⁸ UN Human Rights Council, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns: Lethal Autonomous Robotics and the Protection of Life, A/HRC/23/47, April 9, 2013, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (accessed November 20, 2016), para. 72. 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 ICRC, *Commentary of 1987 on Protocol I*, art. 57, para. 2208 (emphasis added).

²⁰ “Flight of the Drones,” *The Economist*, October 8, 2011, <http://www.economist.com/node/21531433> (accessed November 16) (quoting the US Air Force’s chief scientist, Mark Maybury).

²¹ *Ibid.*

Contention #2: The use of fully autonomous weapons could be limited to specific situations where the weapons would be able to comply with international humanitarian law.

Rebuttal: Narrowly constructed hypothetical cases in which fully autonomous weapons could lawfully be used do not legitimize the weapons because they would likely be used more widely.

Analysis: Some critics, dismissing legal concerns about fully autonomous weapons, contend that their use could be restricted to specific situations where they would be able to conform to the requirements of international humanitarian law. These critics highlight the military utility and low risk to civilians of using the weapons in deserts for attacks on isolated military targets,²² undersea in operations by robotic submarines,²³ in air space for intercepting rockets,²⁴ and for strikes on “nuclear-tipped mobile missile launchers, where millions of lives were at stake.”²⁵ These critics underestimate the threat to civilians once fully autonomous weapons enter military arsenals.

One can almost always describe a hypothetical situation where use of a widely condemned weapon could arguably comply with the general rules of international humanitarian law. Before the adoption of the 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. Once weapons are produced and stockpiled, however, their use is rarely limited to such narrowly constructed scenarios. The widespread use of cluster munitions in populated areas, such as in Iraq in 2003 and Lebanon in 2006, exemplify the reality of this problem.²⁶ Such theoretical possibilities do

²² Kenneth Anderson, Daniel Reisner, and Matthew Waxman, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*, vol. 90 (2014), p. 406.

²³ John Lewis, “The Case for Regulating Fully Autonomous Weapons,” *Yale Law Journal*, vol. 124 (2015), p. 1315.

²⁴ Anderson, Reisner, and Waxman, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*, p. 406.

²⁵ Paul Scharre, “Reflections on the Chatham House Autonomy Conference,” *Lawfare* (blog), March 3, 2014, <http://www.lawfareblog.com/2014/03/guest-post-reflections-on-the-chatham-house-autonomy-conference/> (accessed November 20, 2016).

²⁶ Human Rights Watch, *Off Target: The Conduct of War and Civilian Casualties in Iraq*, December 2003, <https://www.hrw.org/reports/2003/usa1203/usa1203.pdf>, pp. 54-63; Human Rights Watch, *Flooding South Lebanon: Israel’s Use of Cluster Munitions in Lebanon in July and August 2006*, February 2008, <https://www.hrw.org/sites/default/files/reports/lebanono208webwcover.pdf>, pp. 42-44.

not, therefore, legitimize weapons, including fully autonomous ones, that pose significant humanitarian risks when used in less exceptional situations.

Contention #3: Concerns that no one could be held to account for attacks by fully autonomous weapons are of limited importance or could be adequately addressed through existing law.

Rebuttal: Insurmountable legal and practical obstacles would prevent holding anyone responsible for unlawful harms caused by fully autonomous weapons.

Analysis: Some critics argue that the question of accountability for the actions of fully autonomous weapons should not be part of the debate on fully autonomous weapons at all. It would be a mistake to “sacrifice real-world gains consisting of reduced battlefield harm through machine systems ... simply in order to satisfy an *a priori* principle that there must always be a human to hold accountable.”²⁷ Other 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.”²⁸ Accountability is more than what two critics called an “*a priori* principle,” however, and existing mechanisms for legal accountability are ill suited and inadequate to address the unlawful harms fully autonomous weapons would likely cause. These weapons have the potential to commit unlawful acts for which no one could be held responsible.²⁹

Accountability serves multiple moral, social, and political purposes and is a legal obligation. From a policy perspective, it deters future violations, promotes respect for the law, and provides avenues of redress for victims. Redress can encompass retributive justice, which provides the victims the satisfaction that someone was punished for the harm they endured, and compensatory justice to restore victims to the condition they were in before the harm was inflicted.³⁰ International humanitarian law and international human

²⁷ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 17.

²⁸ Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 277.

²⁹ For a more detailed discussion of the accountability gap associated with fully autonomous weapons, see Human Rights Watch and IHRC, *Mind the Gap*.

³⁰ Dinah Shelton, *Remedies in International Human Rights Law* (Oxford: Oxford University Press, 2005), p. 12.

rights law both require accountability for legal violations. International humanitarian law establishes a duty to prosecute criminal acts committed during armed conflict.³¹ International human rights law establishes the right to a remedy for any abuses of human rights (see Contention #5). The value of accountability has been widely recognized, including by scholars and states.³² Unfortunately, the actions of fully autonomous weapons would likely fall into an accountability gap.

Fully autonomous weapons could not be held responsible for their own unlawful acts. Any crime consists of two elements: an act and a mental state. A fully autonomous weapon could commit a criminal act (such as an act listed as an element of a war crime), but it would lack the mental state (often intent) to make these wrongful actions prosecutable crimes. In addition, a weapon would not fall within the natural person jurisdiction of international courts.³³ Even if such jurisdiction were expanded, fully autonomous weapons could not be punished because they would be machines that could not experience or comprehend the significance of suffering.³⁴ Merely altering the software of a “convicted” robot, unable to internalize moral guilt, would likely leave victims seeking retribution unsatisfied.³⁵

In most cases, humans would also escape accountability for the unlawful acts of fully autonomous weapons. Humans could not be assigned direct responsibility for the wrongful actions of a fully autonomous weapon because fully autonomous weapons by definition would have the capacity to act autonomously and therefore could independently and unforeseeably launch an indiscriminate attack against civilians or those *hors de*

³¹ The Fourth Geneva Convention and its Additional Protocol I oblige states to prosecute “grave breaches,” i.e., war crimes, such as willfully targeting civilians or launching an attack with the knowledge it would be disproportionate. Geneva Convention Relative to the Protection of Civilian Persons in Time of War (Fourth Geneva Convention), adopted August 12, 1949, 75 U.N.T.S. 287, entered into force October 21, 1950, art. 146; Protocol I, arts. 85-86.

³² See, for example, Jack M. Beard, “Autonomous Weapons and Human Responsibilities,” *Georgetown Journal of International Law*, vol. 45 (2014); Kelly Cass, “Autonomous Weapons and Accountability,” *Loyola of Los Angeles Law Review*, vol. 48 (2015); Daniel N. Hammond, “Autonomous Weapons and the Problem of State Accountability,” *Chicago Journal of International Law*, vol. 15 (2015); Statement of Norway, Convention on Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 13, 2016 (“Without accountability, deterring and preventing international crimes becomes all that much harder.”); Statement of Pakistan, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11, 2016 (“If the nature of a weapon renders responsibility for its consequences impossible, its use should be considered unethical and unlawful.”).

³³ Rome Statute of the International Criminal Court (Rome Statute), A/CONF.183/9, July 17, 1998, entered into force July 1, 2002, art. 25 (“The Court shall have jurisdiction over natural persons pursuant to this Statute.”); Updated Statute of the International Criminal Tribunal for the Former Yugoslavia, September 2009, art. 6.

³⁴ Robert Sparrow, “Killer Robots,” *Journal of Applied Philosophy*, vol. 24, no. 1 (2007), p. 72.

³⁵ *Ibid.*

combat. In such situations, the commander would not be directly responsible for the robot's specific actions since he or she did not order them. Similarly, a programmer or manufacturer could not be held directly criminally responsible if he or she did not specifically intend, or could not even foresee, the robot's commission of wrongful acts. These individuals could be held directly responsible for a robot's actions only if they deployed the robot intending to commit a crime, such as willfully killing civilians, or if they designed the robot specifically to commit criminal acts.

Significant obstacles would exist to finding the commander indirectly responsible for fully autonomous weapons under the doctrine of command responsibility. This doctrine holds superiors accountable if they knew or should have known of a subordinate's criminal act and failed to prevent or punish it. The autonomous nature of these robots would make them legally analogous to human soldiers in some ways, and thus it could trigger the doctrine. The theory of command responsibility, however, sets a high bar for accountability. Command responsibility deals with prevention of a crime, not an accident or design defect, and robots would not have the mental state to make their unlawful acts criminal.

Regardless of whether the act amounted to a crime, given that these weapons would be designed to operate independently, a commander would not always have sufficient reason or technological knowledge to anticipate the robot would commit a specific unlawful act. Even if he or she knew of a possible unlawful act, the commander would often be unable to prevent the act, for example, if communications had broken down, the robot acted too fast to be stopped, or reprogramming was too difficult for all but specialists. Furthermore, as noted above, punishing a robot is not possible. In the end, fully autonomous weapons would not fit well into the scheme of criminal liability designed for humans, and their use would create the risk of unlawful acts and significant civilian harm for which no one could be held criminally responsible.

An alternative option would be to try to hold the programmer or manufacturer civilly liable for the unanticipated acts of a fully autonomous weapon. Civil liability can be a useful tool for providing compensation, some deterrence, and a sense of justice for those harmed even if it lacks the social condemnation associated with criminal responsibility. There are, however, significant practical and legal obstacles to holding either the programmer or manufacturer of a fully autonomous weapon civilly liable.

On a practical level, most victims would find suing a programmer or manufacturer difficult because their lawsuits would likely be expensive, time consuming, and dependent on the assistance of experts who could deal with the complex legal and technical issues implicated by the use of fully autonomous weapons.

Legal barriers to civil accountability may be even more imposing than practical ones. The doctrine of sovereign immunity protects governments from suits related to the acquisition or use of weaponry, especially in foreign combat situations.³⁶ For example, the US government is presumptively immune from civil suits.³⁷ Manufacturers contracted by the US military are in turn immune from suit when they design a weapon in accordance with government specifications and without deliberately misleading the military. These manufacturers are also immune from civil claims relating to acts committed during wartime. Even without these rules, a plaintiff would find it challenging to establish in law that a fully autonomous weapon was defective for the purposes of a product liability suit.³⁸

A no-fault compensation scheme would not resolve the accountability gap. Such a scheme would require only proof of harm, not proof of defect.³⁹ Victims would thus be compensated for the harm they experienced from a fully autonomous weapon without having to overcome the evidentiary hurdles related to proving a defect. It is difficult to imagine, however, that many governments would be willing to put such a legal regime into place. Even if they did, compensating victims for harm is different from assigning legal responsibility, which establishes moral blame, provides deterrence and retribution, and

³⁶ See, for example, *Just v. British Columbia*, 2 SCR 1228, 1989 (finding the Canadian government immune from suit regarding its policy decisions); UK Crown Proceedings Act 1947, section 10 (stating that the UK government and soldiers themselves are immune for all actions taken by members of the armed forces on duty); *Jurisdictional Immunities of the State (Germany v. Italy)*, International Court of Justice, Judgment, February 3, 2012 (finding that states are immune even for civil suits relating to serious violations of international law).

³⁷ John Copeland Nagle, "Waiving Sovereign Immunity in an Age of Clear Statement Rules," *Wisconsin Law Review*, vol. 1995, (1995), pp. 776-777.

³⁸ See Human Rights Watch and IHRC, *Mind the Gap*, p. 31.

³⁹ Such no-fault systems are often used when a sometimes highly dangerous product or activity is nevertheless deemed socially valuable; they facilitate employment of the risky but useful product by providing compensation to victims, establishing some predictability, and setting limits on the defendant's costs. This type of no-fault system has been used to compensate people injured by vaccines and proposed for self-driving cars. Kevin Funkhouser, "Paving the Road Ahead: Autonomous Vehicles, Products Liability, and the Need for A New Approach," *Utah Law Review*, no. 1 (2013), pp. 458-459; Julie Goodrich, "Driving Miss Daisy: An Autonomous Chauffeur System," *Houston Law Review*, vol. 51 (2013), p. 284.

recognizes victims as persons who have been wronged. Accountability in this full sense cannot be served by compensation alone.⁴⁰

Contention #4: The Martens Clause would not restrict the use of fully autonomous weapons.

Rebuttal: 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. Concerns under 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 to the Geneva Conventions, the Martens Clause mandates that:

In 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.⁴¹

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.”⁴² They contend that because gaps in the law are rare, the probability that fully autonomous weapon would violate the Martens Clause but not applicable treaty and customary law is therefore “exceptionally low.”⁴³ The lack of specific law on fully autonomous weapons, however, means that the Martens Clause would apply, and the weapons would raise serious concerns under the provision.

⁴⁰ For more information, see Human Rights Watch and IHRC, *Mind the Gap*, p. 36.

⁴¹ Protocol 1, art 1(2). The Martens Clause also appears in the preamble of the Hague Convention of 1899. Convention (II) with Respect to the Laws and Customs of War on Land and its Annex: Regulations concerning the Laws and Customs of War on Land, The Hague, adopted July 29, 1899, entered into force September 4, 1900, pmbl.

⁴² Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 275.

⁴³ *Ibid.*, p. 276.

The key question in determining the relevance of the Martens Clause to fully autonomous weapons is the extent to which such weapons would be “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.”⁴⁴ 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.”⁴⁵ Fully autonomous weapons are rapidly evolving forms of technology, at best only generally covered by existing law.⁴⁶

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.⁴⁷ 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.”⁴⁸ According to this view of the Martens Clause, “[i]n case of doubt, international rules, in particular rules belonging to humanitarian law, must be construed so as to be consonant with general

⁴⁴ *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).

⁴⁵ 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 November 20, 2016), para. 78.

⁴⁶ 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 Contention #1. 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 fully autonomous weapons that launched landmines, cluster munitions, or incendiary 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.

⁴⁷ 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 most of the states that appeared 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); ICRC, *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 November 20, 2016), 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”).

⁴⁸ Cassese, “The Martens Clause,” *European Journal of International Law*, p. 212.

standards of humanity and the demands of public conscience.”⁴⁹ Given the significant doubts about the ability of fully autonomous weapons to conform to the requirements of the law (see Contention #1), the standards of the Martens Clause should at the very least be taken into account when evaluating the weapons’ legality.

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.⁵⁰ As discussed below under Contention #7, fully autonomous weapons would lack human emotions, including compassion. The challenges the weapons would face in meeting international humanitarian law suggest they could not adequately protect civilians. 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. For example, a 2015 international survey of 1,002 individuals from 54 different countries found that 56 percent of respondents opposed the development and use of these weapons.⁵¹ The first reason given for rejecting their development and use, cited by 34 percent of all respondents, was that “humans should always be the one to make life/death decisions.”⁵² A 2013 national survey of Americans found that 68 percent of respondents with a view on the topic opposed the move toward these weapons (48 percent strongly).⁵³ Interestingly, active duty military personnel were among the strongest objectors—73 percent expressed opposition

⁴⁹ 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=EPIL> (accessed November 20, 2016), 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’.”).

⁵⁰ 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 20, 2016), p. 2.

⁵¹ Open Roboethics Initiative, “The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll,” November 9, 2015, http://www.openroboethics.org/wp-content/uploads/2015/11/ORi_LAWS2015.pdf (accessed November 20, 2016), pp. 4, 8.

⁵² Ibid., p. 7.

⁵³ Charli Carpenter, “US Public Opinion on Autonomous Weapons,” June 2013, http://www.duckofminerva.com/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf (accessed November 20, 2016). 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 percent. A discussion of the sampling methods, limitations, and accuracy can be found at <http://yougov.co.uk/publicopinion/methodology/>.

to fully autonomous weapons. These kinds of reactions suggest that fully autonomous weapons would 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 preemptively banned the transfer and use of laser weapons whose sole or partial purpose is to cause permanent blindness.⁵⁴ 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.⁵⁵ They largely agreed that “[blinding lasers] would run counter to the requirements of established custom, humanity, and public conscience.”⁵⁶ A shared horror at the prospect of blinding weapons ultimately helped tip the scales toward a prohibition, even without consensus that such weapons were unlawful under the core principles of international humanitarian law.⁵⁷ The Blinding Lasers Protocol set an international precedent for preemptively banning weapons based, at least in part, on the Martens Clause.⁵⁸ Invoking the clause in the context of fully autonomous weapons would be equally appropriate.

⁵⁴ Human Rights Watch and IHRC, “Precedent for Preemption,” pp. 3-7. See also 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; CCW Protocol on Blinding Lasers (CCW Protocol IV), adopted October 13, 1995, entered into force July 30, 1998, art. 1.

⁵⁵ 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; 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. 344-346.

⁵⁶ 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*, pp. 344-346. Others doubted their ability to field such weapons, notwithstanding possible military utility, because of public opinion. *Ibid.*, p. 345.

⁵⁷ This reaction is suggested by the comments of the participating experts in the ICRC meetings. For example, one participant stated 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.

⁵⁸ Akerson, “The Illegality of Offensive Lethal Autonomy,” p. 96.

Contention #5: International humanitarian law is the only relevant body of law under which to assess fully autonomous weapons because they would be tools of armed conflict.

Rebuttal: An assessment of fully autonomous weapons must consider their ability to comply with all bodies of international law, including international human rights law, because the weapons could be used outside of armed conflict situations. Fully autonomous weapons could violate the right to life, the right to a remedy, and the principle of dignity, each of which is guaranteed by international human rights law.

Analysis: Discussions about fully autonomous weapons have largely focused on their use in armed conflict and their legality under international humanitarian law (see Contention #1). Most of the diplomatic debate about the weapons has taken place in the international humanitarian law forum of the CCW. While states have touched on the human rights implications of fully autonomous weapons in CCW meetings and in the Human Rights Council, the weapons' likely use beyond the battlefield has often been ignored.⁵⁹ Human rights law, which applies during peace and war, would be relevant to all circumstances in which fully autonomous weapons might be used, and thus should receive greater attention.⁶⁰

Once developed, fully autonomous weapons could be adapted to a range of non-conflict contexts that can be grouped under the heading of law enforcement. Local police officers could potentially use such weapons in crime fighting, the management of public protests, riot control, and other efforts to maintain law and order. States could also utilize the weapons in counter-terrorism efforts falling short of an armed conflict as defined by international humanitarian law. The use of fully autonomous weapons in a law enforcement context would trigger the application of international human rights law.

Fully autonomous weapons would have the potential to contravene the right to life, which is codified in Article 6 of the International Covenant on Civil and Political Rights (ICCPR):

⁵⁹ Christof Heyns, "Human Rights and the Use of Autonomous Weapons Systems (AWS) during Domestic Law Enforcement," *Human Rights Quarterly*, vol. 38, p. 351, n. 2.

⁶⁰ For a more detailed discussion of the human rights implications of fully autonomous weapons, see Human Rights Watch and IHRC, *Shaking the Foundations*.

“Every human being has the inherent right to life. This right shall be protected by law.”⁶¹ The Human Rights Committee, the ICCPR’s treaty body, describes it as “the supreme right” because it is a prerequisite for all other rights.⁶² It is non-derogable even in public emergencies that threaten the existence of a nation. The right to life prohibits arbitrary killing. The ICCPR states, “No one shall be arbitrarily deprived of his life.”⁶³

The right to life constrains the application of force in law enforcement situations, including those in which fully autonomous weapons could be deployed.⁶⁴ In its General Comment No. 6, the Human Rights Committee highlights the duty of states to prevent arbitrary killings by their security forces.⁶⁵ Killing is only lawful if it meets three cumulative requirements for when and how much force may be used: it must be necessary to protect human life, constitute a last resort, and be applied in a manner proportionate to the threat. Fully autonomous weapons would face significant challenges in meeting the criteria circumscribing lawful force because the criteria require qualitative assessments of specific situations. These robots could not be programed in advance to assess every situation because there are infinite possible scenarios, a large number of which could not be anticipated. According to many roboticists, it is also highly unlikely in the foreseeable future that robots could be developed to have certain human qualities, such as judgment and the ability to identify with humans, that facilitate compliance with the three criteria.⁶⁶ A fully autonomous weapon’s misinterpretation of the appropriateness of using force could trigger an arbitrary killing in violation of the right to life.

⁶¹ International Covenant on Civil and Political Rights (ICCPR), adopted December 16, 1966, G.A. Res. 2200A (XXI), 21 U.N. GAOR Supp. (No. 16) at 52, U.N. Doc. A/6316 (1966), 999 U.N.T.S. 171, entered into force March 23, 1976, art. 6.

⁶² UN Human Rights Committee, General Comment No. 6, Right to Life, U.N. Doc. HRI/GEN/1/Rev.1 at 6 (1994), para. 1. See also Manfred Nowak, *U.N. Covenant on Civil and Political Rights: CCPR Commentary* (Arlington, VA: N.P. Engel, 2005), p. 104.

⁶³ ICCPR, art. 6(1).

⁶⁴ See Human Rights Watch and IHRC, *Shaking the Foundations*, pp. 8-14.

⁶⁵ UN Human Rights Committee, General Comment No. 6, para. 3.

⁶⁶ See Marcello Guarini and Paul Bello, “Robotic Warfare: Some Challenges in Moving from Noncivilian to Civilian Theaters,” in *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. Patrick Lin, Keith Abney, and George A. Bekey (Cambridge, MA: Massachusetts Institute of Technology, 2012), p. 138 (“A system without emotion ... could not predict the emotions or action of others based on its own states because it has no emotional states.”); Noel Sharkey, “Killing Made Easy: From Joysticks to Politics,” in *Robot Ethics*, eds. Lin, Abney, and Bekey, p. 118 (“Humans understand one another in a way that machines cannot. Cues can be very subtle, and there are an infinite number of circumstances where lethal force is inappropriate.”).

As a non-derogable right, the right to life continues to apply during armed conflict.⁶⁷ In wartime, arbitrary killing refers to unlawful killing under international humanitarian law. In his authoritative commentary on the ICCPR, Manfred Nowak, former UN special rapporteur on torture, defines arbitrary killings in armed conflict as “those that contradict the humanitarian laws of war.”⁶⁸ As has been shown under Contention #1, there are serious doubts as to whether fully autonomous weapons could ever comply with rules of distinction and proportionality. Fully autonomous weapons would have the potential to kill arbitrarily and thus violate the right that underlies all others, the right to life.

The use of fully autonomous weapons also threatens to contravene the right to a remedy. The Universal Declaration of Human Rights (UDHR) lays out the right, and Article 2(3) of the ICCPR requires states parties to “ensure that any person whose rights or freedoms ... are violated shall have an effective remedy.”⁶⁹ The right to a remedy requires states to ensure individual accountability. It includes the duty to prosecute individuals for serious violations of human rights law and punish individuals who are found guilty.⁷⁰ International law mandates accountability in order to deter future unlawful acts and punish past ones, which in turn recognizes victims’ suffering. It is unlikely, however, that meaningful accountability for the actions of a fully autonomous weapon would be possible (see Contention #3).

Fully autonomous weapons could also violate the principle of dignity, which is recognized in the opening words of the UDHR.⁷¹ As inanimate machines, fully autonomous weapons

⁶⁷ In its advisory opinion on nuclear weapons, the International Court of Justice found: “In principle, the right not arbitrarily to be deprived of one’s life applies also in hostilities.” International Court of Justice, *Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons*, para. 25. See also Nowak, *U.N. Covenant on Civil and Political Rights*, p. 108 (“Arbitrary killings in the course of armed conflicts permissible under international law and civil wars also represent a violation of the right to life.”); UN Human Rights Committee, General Comment No. 31, *The Nature of the General Legal Obligation Imposed on States Parties to the Covenant* (Eightieth Session, 2004), U.N. Doc. CCPR/C/21/Rev.1/Add.13 (2004), para. 11 (The ICCPR “applies also in situations of armed conflict to which the rules of international humanitarian law are applicable.”).

⁶⁸ Nowak, *U.N. Covenant on Civil and Political Rights*, p. 108, n. 29. Given that states “have the supreme duty to prevent wars,” killings in the course of a war that violates the UN Charter would also violate the right to life. *Ibid.*, p. 108. See also UN Human Rights Committee, General Comment No. 6, para. 2.

⁶⁹ Universal Declaration of Human Rights (UDHR), adopted December 10, 1948, G.A. Res. 217A(III), U.N. Doc. A/810 at 71 (1948), art. 8 (“Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by the constitution or by law.”); ICCPR, art. 2(3).

⁷⁰ See UN Human Rights Committee, General Comment No. 31, paras. 15, 18; Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law (2005 Basic Principles and Guidelines), adopted December 16, 2005, G.A. Res. 60/47, art. 4.

⁷¹ UDHR, pmb., para. 1 (“recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world.”).

could truly comprehend neither the value of individual life nor the significance of its loss, and thus should not be allowed to make life-and-death decisions (see Contention #6).

Non-Legal Dangers

Contention #6: Moral concerns about fully autonomous weapons either are irrelevant or could be overcome.

Rebuttal: A variety of actors have raised strong and persuasive moral objections to fully autonomous weapons, most notably related to the weapons' lack of judgment and empathy, threat to dignity, and absence of moral agency.

Analysis: Some critics dismiss questions about the morality of fully autonomous weapons as irrelevant. They say the appropriateness of fully autonomous weapons is a legal and technical matter as opposed to a moral one. One critic writes that the “key issue remains whether or not a particular weapon system can be operated in compliance with IHL rules and obligations, not the presence or absence of a human moral agent.”⁷² At least one other critic argues that morality would not be an issue because robots could be programmed to act ethically and could thus constitute moral agents.⁷³ Concerns about the morality of fully autonomous weapons, however, are foundational and far reaching.

A variety of actors have raised strong moral and ethical concerns about the use of fully autonomous weapons. The moral indignation expressed by states, UN special rapporteurs, Nobel peace laureates, religious leaders, and the public shows that the question of whether fully autonomous weapons should ever be used goes beyond the law. Several states have argued that there is a moral duty to maintain human control.⁷⁴ A 2015 paper from the Holy See, which has presented the most in-depth discussion of the ethical objections to fully autonomous weapons, explained, “It is fundamentally immoral to utilize a weapon the behavior of which we cannot completely control.”⁷⁵ The previous year, Chile stated that significant human control over weapons is an “ethical imperative” rather than a

⁷² Beard, “Autonomous Weapons and Human Responsibilities,” *Georgetown Journal of International Law*, p. 640.

⁷³ Arkin, *Governing Lethal Behavior in Autonomous Robots*, p. 127.

⁷⁴ At least 16 states raised ethical concerns at the CCW Meetings of Experts on Lethal Autonomous Weapons Systems in 2014, 2015, and 2016.

⁷⁵ Statement of the Holy See, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 16, 2015, p. 8.

technological problem.⁷⁶ According to then UN Special Rapporteur on Extrajudicial Killing Christof Heyns, whether fully autonomous weapons are morally unacceptable “is an overriding consideration” and “no other consideration can justify the deployment of [fully autonomous weapons], no matter the level of technical competence at which they operate.”⁷⁷ Heyns and Maina Kiai, special rapporteur on the rights to freedom of peaceful assembly and of association, have both called for a ban on these weapons.⁷⁸ Nobel Peace Prize laureates have stressed the need to outline “the moral and legal perils of creating killer robots and call[ed] for public discourse before it is too late.”⁷⁹ According to Nobel Laureate Jody Williams, who is a member of the Campaign to Stop Killer Robots, “Where is humanity going if some people think it’s OK to cede the power of life and death of humans over to a machine?”⁸⁰ A religious leaders’ interfaith declaration calling for a ban highlighted moral and ethical concerns, stating that “[r]obotic warfare is an affront to human dignity and to the sacredness of life.”⁸¹ Research surveys conducted in the United States and internationally have shown that these moral concerns are shared among populations around the world.⁸²

For those concerned with the moral issues raised by fully autonomous weapons, no technological improvements can solve the fundamental problem of delegating a life-and-death decision to a machine. Morality-based arguments have focused on three core issues: the lack of human qualities necessary to make a moral decision, the threat to

⁷⁶ Statement of Chile, CCW Meeting of States Parties, Geneva, November 13-14, 2014.

⁷⁷ UN Human Rights Council, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns: Lethal Autonomous Robotics and the Protection of Life, para. 93.

⁷⁸ Joint Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association and the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions on the Proper Management of Assemblies to the Human Rights Council, A/HRC/31/66, February 4, 2016, para. 67(f).

⁷⁹ “Nobel Peace Laureates Call for Preemptive Ban on Killer Robots,” May 12, 2014, <http://nobelwomensinitiative.org/nobel-peace-laureates-call-for-preemptive-ban-on-killer-robots/> (accessed November 21, 2016).

⁸⁰ John Thornhill, “Military Killer Robots Create a Moral Dilemma,” *Financial Times*, April 25, 2016, <https://www.ft.com/content/8deae2c2-088d-11e6-a623-b84do6a39ec2> (accessed October 20, 2016) (quoting Jody Williams).

⁸¹ Pax Christi International, “Interfaith Declaration in Support of a Ban on Fully Autonomous Weapons,” http://www.paxchristi.net/sites/default/files/interfaith_declaration.pdf (accessed November 21, 2016).

⁸² See, for example, Charli Carpenter, “Who’s Afraid of Killing Robots? (and Why),” *Washington Post*, May 30, 2014, <https://www.washingtonpost.com/news/monkey-cage/wp/2014/05/30/whos-afraid-of-killer-robots-and-why/> (accessed November 21, 2016) (noting, “According to respondents, the key human quality machines would presumably lack would be a moral conscience. Respondents repeatedly characterized judgment, empathy and moral reasoning as uniquely human traits.”). See also Open Roboethics Initiative, “The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll,” p. 7 (reporting that 34 percent of respondents gave “Humans should always be the one to make life/death decisions” as their main reason for rejecting the development and use of fully autonomous weapons).

human dignity, and the absence of moral agency.

Any killing orchestrated by a machine is arguably inherently wrong since machines are unable to exercise human judgment and compassion. Because of the high value of human life, a decision to take a life deliberately is extremely grave. As humans are endowed with reason and intellect, they are uniquely qualified to make the moral decision to apply force in any particular situation. Humans possess “prudential judgment,” the ability to apply broad principles to particular situations, interpreting and giving a “spirit” to laws rather than blindly applying an algorithm.⁸³ No robot, however much information it can process, possesses prudential judgment in the same way that humans do. In addition, while humans in some way internalize the cost of any life that they choose to take, machines do not.⁸⁴ “Decisions over life and death in armed conflict may require compassion and intuition,” which humans, not robots, possess.⁸⁵ This allows for human empathy to act as a check on killing, but only when humans are making the relevant decisions.

Fully autonomous weapons are also morally problematic because they threaten the principle of human dignity. The opening words of the Universal Declaration of Human Rights assert that “recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world.”⁸⁶ (For other human rights arguments, see Contention #5.) In ascribing inherent dignity to all human beings, the UDHR implies that everyone has worth that deserves respect.⁸⁷ Fully autonomous weapons, as inanimate machines, could comprehend neither the value of individual life nor the significance of its loss. Allowing them to make determinations to take life away would thus conflict with the principle of dignity. Indeed, as one author notes, the “value of human life may be diminished if

⁸³ Holy See, “The Use of Lethal Autonomous Weapon Systems: Ethical Questions,” CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 13, 2015, [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/4D28AF2B8BBBECEDC1257E290046B73F/\\$file/2015_LAWS_MX_Holy+See.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/4D28AF2B8BBBECEDC1257E290046B73F/$file/2015_LAWS_MX_Holy+See.pdf) (accessed November 21, 2016) (“Prudential judgement cannot be put into algorithms.”).

⁸⁴ UN Human Rights Council, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns: Lethal Autonomous Robotics and the Protection of Life, para. 94.

⁸⁵ *Ibid.*, para. 55.

⁸⁶ UDHR, pmb., para. 1. The Oxford English Dictionary defines dignity as “the quality of being worthy or honourable; worthiness, worth, nobleness, excellence.” Oxford English Dictionary online, “Dignity.”

⁸⁷ Jack Donnelly, “Human Dignity and Human Rights,” in Swiss Initiative to Commemorate the 60th Anniversary of the UDHR, *Protecting Dignity: Agenda for Human Rights*, June 2009, <https://www.scribd.com/document/200255016/HUMAN-DIGNITY-AND-HUMAN-RIGHTS> (accessed November 21, 2016), p. 10.

machines are in a position to make essentially independent decisions about who should be killed in armed conflict.”⁸⁸ Then Special Rapporteur on Extrajudicial Killing Christof Heyns, in his 2013 report to the Human Rights Council, stated: “[D]elegating this process dehumanizes armed conflict even further and precludes a moment of deliberation in those cases where it may be feasible. Machines lack morality and mortality, and should as a result not have life and death powers over humans.”⁸⁹

Fully autonomous weapons raise further concerns under the umbrella of moral agency. According to one roboticist, agency is not an issue: such machines could be programmed to operate on the basis of “ethical” algorithms that would transform an autonomous robot into a “moral machine” and in this way into an “autonomous moral agent.”⁹⁰ An “ethical governor” would automate moral decision making at the targeting and firing stages.⁹¹ This argument is unpersuasive for two reasons, however. First, it is extremely unlikely that such a protocol will ever be designed.⁹² Second, and more fundamentally, “the problem of moral agency is not solved by giving autonomous weapon systems artificial moral judgment, even if such a capacity were technologically possible.”⁹³ “Fully ethical agents” are endowed with “consciousness, intentionality and free will.”⁹⁴ Fully autonomous weapons, by contrast, would act according to algorithms and thus would not be moral agents. Fully ethical agents “can be held accountable for their actions—in the moral sense, they can be at fault—precisely because their decisions are in some rich sense up to them.”⁹⁵ Fully autonomous weapons, on the other hand, would be incapable of assuming

⁸⁸ Roni Elias, “Facing the Brave New World of Killer Robots,” *Indonesian Journal of International & Comparative Law*, vol. 3 (2016), p. 115.

⁸⁹ UN Human Rights Council, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns: Lethal Autonomous Robotics and the Protection of Life, para. 94.

⁹⁰ Arkin, *Governing Lethal Behavior in Autonomous Robots*, p. 127. For further discussion of Arkin’s position, see Wendell Wallach and Colin Allen, *Moral Machines* (Oxford: Oxford University Press, 2008), pp. 171-172.

⁹¹ Arkin, *Governing Lethal Behavior in Autonomous Robots*, p. 127.

⁹² Duncan Purves, Ryan Jenkins, and Bradley J. Strawser, “Autonomous Machines, Moral Judgment, and Acting for the Right Reasons,” *Ethical Theory and Moral Practice*, vol. 18, no. 4 (2015), pp. 855-858 (discussing the impossibility of codifying moral judgment); Elias, “Facing the Brave New World of Killer Robots,” *Indonesian Journal of International & Comparative Law*, p. 122. See also Heather M. Roff, “The Strategic Robot Problem: Lethal Autonomous Weapons in War,” *Journal of Military Ethics*, vol. 13, no. 3 (2014), pp. 213-215; Matthias Englert, Sandra Siebert, and Martin Ziegler, “Logical Limitations to Machine Ethics with Consequences to Lethal Autonomous Weapon,” IANUS, Technische Universität Darmstadt (2014) (discussing the impossibility of programming a robot to ensure that it is consistently able to recognize the moral decision in any given situation).

⁹³ Elias, “Facing the Brave New World of Killer Robots,” *Indonesian Journal of International & Comparative Law*, p. 122.

⁹⁴ James H. Moor, “The Nature, Importance and Difficulty of Machine Ethics,” *IEEE Intelligent Systems* (2006), p.20.

⁹⁵ Anthony Beavers, “Moral Machines and the Threat of Ethical Nihilism,” in *Robot Ethics*, eds. Lin, Abney, and Bekey, p. 6.

moral responsibility for their actions and thus could not meet the threshold of moral agency that is required for the taking of human life.⁹⁶

Technological improvements could not overcome such moral objections to fully autonomous weapons. As one expert wrote, "The authority to decide to initiate the use of lethal force ... must remain the responsibility of a human with the duty to make a considered and informed decision before taking human lives."⁹⁷

Contention #7: Fully autonomous weapons would not be negatively influenced by human emotions.

Rebuttal: 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.⁹⁸ While such observations have some merit, the role in warfare of other human emotions can in fact increase 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 their fellow man. A resistance so strong that, in many circumstances, soldiers on the battlefield will die before they can overcome it."⁹⁹ Another author writes,

⁹⁶ Beavers, "Moral Machines and the Threat of Ethical Nihilism." See also Samir Chopra and Laurence F. White, *A Legal Theory for Autonomous Artificial Agents* (University of Michigan Press 2011); Wallach and Allen, *Moral Machines*.

⁹⁷ Peter Asaro, "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making," *International Review of the Red Cross*, vol. 94, no. 886 (2012), p. 686.

⁹⁸ 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 November 21, 2016), pp. 6-7.

⁹⁹ 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.

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.¹⁰⁰

Studies of soldiers' conduct in past conflicts provide evidence to support these conclusions.¹⁰¹ 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 armed conflict, 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 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."¹⁰²

Due to their lack of emotions or a conscience, 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

¹⁰⁰ Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Farnham: Ashgate Publishing Limited, 2009), p. 130.

¹⁰¹ 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").

¹⁰² Krishnan, *Killer Robots*, p. 130.

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

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

Contention #8: Military advantages would be lost with a preemptive ban on fully autonomous weapons.

Rebuttal: Many potential benefits of fully autonomous weapons either could be achieved by using alternative systems or would create unjustifiable risks.

Analysis: Critics argue that a preemptive ban on fully autonomous weapons would mean forgoing the technology's touted military advantages. According to these critics, fully autonomous weapons could have many benefits. Fully autonomous weapons could operate with greater precision than other systems.¹⁰³ The weapons could replace soldiers in the field and thus protect their lives.¹⁰⁴ Fully autonomous weapons could process data and operate at greater speed than those controlled by humans at the targeting and/or engagement stages.¹⁰⁵ They could also operate without a line of communication after deployment.¹⁰⁶ Finally, fully autonomous weapons could be deployed on a greater scale and at a lower cost than weapons systems requiring human control.¹⁰⁷ These characteristics, however, are not unique to fully autonomous weapons and present their own risks.

Other weapons provide some of the same benefits as fully autonomous weapons. For example, semi-autonomous weapons, too, have the potential for precision. They can track

¹⁰³ Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems," *Jean Perkins Task Force on National Security and Law*, p. 2.

¹⁰⁴ Hammond, "Autonomous Weapons and the Problem of State Accountability," *Chicago Journal of International Law*, p. 661.

¹⁰⁵ *Ibid.*, p. 660.

¹⁰⁶ Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems," *Jean Perkins Task Force on National Security and Law*, p. 5; Hammond, "Autonomous Weapons and the Problem of State Accountability," *Chicago Journal of International Law*, pp. 660-661.

¹⁰⁷ Hammond, "Autonomous Weapons and the Problem of State Accountability," *Chicago Journal of International Law*, p. 660.

targets with comparable technology to that in future fully autonomous weapons. Indeed, existing semi-autonomous weapon systems have already incorporated autonomous features designed to increase the precision of attacks.¹⁰⁸ Unlike their fully autonomous counterparts, however, these systems keep a human in the loop on decisions to fire.

In addition, although fully autonomous weapons could reduce military casualties by replacing human troops on the battlefield, semi-autonomous weapons already do that. The use of semi-autonomous weapons involves human control over the use of force, but it does not require a human presence on the ground so operators can stay safe at a remote location. Semi-autonomous weapons, notably armed drones, have raised many concerns that should be addressed, but their problems relate more to how they are used than to the nature of their technology. Fully autonomous weapons, by contrast, present dangers no matter how they are used because humans are no longer making firing decisions.

In many situations that require speed, such as missile defense, automatic systems could eliminate threats as effectively as and more predictably than fully autonomous systems. While automation and autonomy are different ends of the same spectrum, automatic weapons operate in a more structured environment and “carr[y] out a pre-programmed sequence of operations.”¹⁰⁹

Because fully autonomous weapons would have the power to make complex determinations in less structured environments, their speed could lead armed conflicts to spiral rapidly out of control. In arguing that fully autonomous weapons could become a necessity for states seeking to keep up with their adversaries, two critics of a ban on fully autonomous weapons 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.”¹¹⁰ Regardless of the speed of fully autonomous weapons, their ability to operate without a line of

¹⁰⁸ “Preparing for the Future of Artificial Intelligence,” Executive Office of the President, National Science and Technology Council Committee on Technology, October 2016, https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (accessed November 21, 2016), p. 1 (“The United States has incorporated autonomy in certain weapon systems for decades, allowing for greater precision in the use of weapons and safer, more humane military operations.”).

¹⁰⁹ Noel Sharkey, “Automating Warfare: Lessons Learned from the Drones,” *Journal of Law, Information & Science* (2011), <http://www.jlisjournal.org/abstracts/sharkey.21.2.html>, (accessed November 21, 2016), p. EAP 2.

¹¹⁰ Schmitt and Thurnher, “Out of the Loop,” *Harvard National Security Journal*, p. 238.

communication after deployment is problematic because the weapons could make poor, independent choices about the use of force absent the potential of a human override.

Since fully autonomous weapons could operate at high speeds and without human control, their actions would also not be tempered by human understanding of political, socioeconomic, environmental, and humanitarian risks at the moment they engage. They would thus have the potential to trigger a range of unintended consequences, many of which could fundamentally alter relations between states or the nature of ongoing conflicts.

Given that countries would not want to fall behind in potentially advantageous military technology, the development of these revolutionary weapons would likely lead to an arms race. Indeed, some senior military officials have already expressed concerns about advancements in autonomous weapons technology in other states, emphasizing the need to maintain dominance in artificial intelligence capabilities.¹¹¹ High-tech militaries might have an edge in the early stages of these weapons' development, but experts predict that as costs go down and the technology proliferates, the weapons will become mass produced. An open letter signed by more than 3,000 artificial intelligence and robotics experts states:

If any major military power pushes ahead with AI [artificial intelligence] weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce.¹¹²

An arms race in fully autonomous weapons technology would carry significant risks. The rapidly growing number of fully autonomous weapons could heighten the possibility of major conflict. If fully autonomous weapons operated collectively, such as in swarms, one

¹¹¹ Patrick Tucker, "The Pentagon is Nervous about Russian and Chinese Killer Robots," *Defense One*, December 14, 2015, <http://www.defenseone.com/threats/2015/12/pentagon-nervous-about-russian-and-chinese-killer-robots/124465/?oref=DefenseOneFB&&> (accessed November 21, 2016).

¹¹² Future of Life Institute, "Autonomous Weapons: An Open Letter from AI & Robotics Researchers" opened July 28, 2015, <http://futureoflife.org/open-letter-autonomous-weapons> (accessed November 21, 2016).

weapon's malfunction could trigger a massive military action followed by a response in kind.¹¹³ Moreover, in order to keep up with their enemies, states would have incentive to use substandard fully autonomous weapons with untested or outdated features, increasing the risk of potentially catastrophic errors. While fully autonomous weapons might create an immediate military advantage for some states, they should recognize that it would be short lived once the technology began to proliferate. Ultimately, the financial and human costs of developing such technology would leave each state worse off, and thus they argue for a preemptive ban.

¹¹³ Frank Sauer, International Committee for Robot Arms Control (ICRAC), "ICRAC Second Statement on Security to the 2016 UN CCW Expert Meeting," April 15, 2016, <http://icrac.net/2016/04/icrac-second-statement-on-security-to-the-2016-un-ccw-expert-meeting/> (accessed November 21, 2016).

II. Arguments for a Preemptive Prohibition on Fully Autonomous Weapons

The dangers of fully autonomous weapons demand that states take action to preemptively ban their development, production, and use. Critics propose relying on existing law, weapons reviews, regulation, or requirements of human control, but a ban is the only option that would address all of the weapons' problems. The international community should not wait to take action because the genie will soon be out of the bottle. Precedent shows that a ban would be achievable and effective.

Advantages of a Ban

Contention #9: A new international instrument is unnecessary because existing international humanitarian law will suffice.

Rebuttal: A new treaty would help clarify existing international humanitarian law and would address the 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 principles of international law are sufficient to circumscribe the use of these weapons.”¹¹⁴ 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.”¹¹⁵ 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],” but he too concludes that a new legal instrument would be unnecessary.¹¹⁶ Existing international humanitarian law, however, was not intended to and

¹¹⁴ Lewis, “The Case for Regulating Fully Autonomous Weapons,” *Yale Law Journal*, p. 1325. See also Cass, “Autonomous Weapons and Accountability,” *Loyola of Los Angeles Law Review*, pp. 1039-1040.

¹¹⁵ Schmitt and Thurnher, “‘Out of the Loop,’” *Harvard National Security Journal*, p. 232.

¹¹⁶ Benjamin Kastan, “Autonomous Weapons Systems: A Coming Legal ‘Singularity?’” *University of Illinois Journal of Law, Technology, and Policy* (Spring 2013), p. 45.

cannot adequately address the issues raised by this revolutionary type of weapon. Therefore, it should be supplemented with a new treaty establishing a ban.

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 (see Contention #10). Finally, many states that did not join the new treaty would still be apt to abide by its ban because of the stigma associated with the weapons.

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 (see Contention #8).

Finally, new law could address concerns about an accountability gap (see Contention #3). A treaty that banned fully autonomous weapons under any circumstances could require that anyone violating that rule be held responsible for the 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.

Contention #10: Reviews of new weapons systems can address the dangers of fully autonomous weapons.

Rebuttal: Weapons reviews are not universal, consistent, or rigorously conducted, and they fail to address the implications of weapons outside of an armed conflict context. A ban would resolve these shortcomings in the case of fully autonomous weapons.

Analysis: Some critics argue that conducting weapons reviews on fully autonomous weapons would sufficiently regulate the weapons. Weapons reviews assess the legality of the future use of a new weapon during its design, development, and acquisition phases. They are sometimes called “Article 36 reviews” because they are required under Article 36 of Additional Protocol I to the Geneva Conventions. The article states:

In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.¹¹⁷

Critics have argued, including during CCW debates, that there is no need for a ban because any fully autonomous weapon that would violate international law would fail a weapons review and thus not be developed or used.¹¹⁸ Not all governments, however, conduct weapons reviews, those that do follow varying standards, and reviews are often too narrow in scope sufficiently to address every danger posed by fully autonomous weapons.

¹¹⁷ Protocol I, art. 36.

¹¹⁸ See, for example, William Boothby, Expert Presentation at CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 13-17, 2015, [http://www.unog.ch/80256EE600585943/\(httpPages\)/6CE049BE22EC75A2C1257C8D00513E26?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/6CE049BE22EC75A2C1257C8D00513E26?OpenDocument) (accessed November 21, 2016) (saying, “the proper answer has to be ensuring that states properly review weapons”); Jai Galliot, Expert Presentation at the CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11-15, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument) (accessed November 21, 2016) (“existing international law and weapons review procedures serve an adequate regulatory function”); Anderson, Reisner, and Waxman, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*, p. 398.

Proposals to address the shortcomings of weapons reviews should be considered in a separate forum to avoid distracting from discussions about fully autonomous weapons. Currently, fewer than 30 states are known to have national review processes in place.¹¹⁹ Not all states are party to Additional Protocol I, and it is debated whether weapons reviews are required under customary international law.¹²⁰ The lack of universal practice means that it is possible that some states could develop or acquire fully autonomous weapons without first reviewing the legality of the weapons at all.

Even if weapons reviews were conducted by every state, leaving decisions about whether or not to develop weapons to individual states is bound to lead to inconsistent outcomes. The complexity of fully autonomous weapons, which would require review of both hardware and software components, would exacerbate such inconsistencies.¹²¹ In addition, there is no internationally mandated monitoring to ensure that all states conduct reviews and adhere to the results.¹²² There is also limited capability for outside monitoring, including by civil society, because of the general lack of transparency in weapons reviews processes.¹²³ States are not obliged to release their reviews, and none are known to have disclosed information about a review that rejected a proposed weapon.¹²⁴

Without the external pressure generated by monitoring, states have few incentives to conduct rigorous reviews of weapons. Just as there are no publicized cases of the rejection of a weapon, there are also no known examples of states stopping the development or production of a weapon because it failed a legal review.¹²⁵ The expense of conducting the

¹¹⁹ Article 36, “Article 36 Reviews and Addressing Lethal Autonomous Weapons Systems: Briefing Paper for Delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).” April 2016, <http://www.article36.org/wp-content/uploads/2016/04/LAWS-and-A36.pdf> (accessed November 21, 2016), p. 3.

¹²⁰ ICRC, *A Guide to the Legal Review of New Weapons*, p. 4. See also Cass, “Autonomous Weapons and Accountability,” *Loyola of Los Angeles Law Review*, p. 1041.

¹²¹ Vincent Boulanin, “Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapon Systems,” *SIPRI Insights on Peace and Security*, no. 2015/1 (November 2015), <https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf> (accessed November 21, 2016), p. 30.

¹²² Protocol I, art. 36.

¹²³ See, for example, Statements of Canada, Finland, Sweden, and Zambia, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11-15, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAF2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAF2?OpenDocument) (accessed November 21, 2016).

¹²⁴ Article 36, “Article 36 Reviews,” p. 2.

¹²⁵ *Ibid.*

kind of complex reviews necessary for fully autonomous weapons would provide a further disincentive to doing rigorous testing.

Regardless of the effectiveness of the weapons reviews, the basic goal, as evidenced by Article 36's reference to "warfare," is to ensure compliance with international law in the context of armed conflict. The ICRC's guide to weapons reviews reflects this framework, noting that "[a]ssessing the legality of new weapons contributes to ensuring that a State's armed forces are capable of conducting hostilities in accordance with its international obligations."¹²⁶

This framework does not address the human rights and ethical implications of the use of weapons. Fully autonomous weapons could independently contravene human rights law because of their potential use outside of armed conflict in domestic law enforcement situations (see Contention #5).¹²⁷ Because they would use force without meaningful human control, such weapons raise serious ethical concerns (see Contention #6). Neither of these risks would be taken into account in a military weapons review.¹²⁸

Acknowledging the problems with existing weapons reviews, some states have called for improvements.¹²⁹ For example, at the 2016 CCW Meeting of Experts on Lethal Autonomous Weapons Systems, the United States proposed that CCW states parties produce "a non-legally binding outcome document that describes a comprehensive weapons review

¹²⁶ ICRC, *A Guide to the Legal Review of New Weapons*, p. 1.

¹²⁷ See generally Human Rights Watch and IHRC, *Shaking the Foundations*.

¹²⁸ Article 36, "Article 36 Reviews," p. 2. See also Miriam Struyk, PAX, "Transparency is Not Enough," CCW side event presentation, April 13, 2016, <https://wapenfeiten.files.wordpress.com/2016/04/presentation-pax-side-event-april-2016.pdf> (accessed November 21, 2016); Statement of Mines Action Canada, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 13, 2016, <https://stopkillerrobots.ca/2016/04/13/mines-action-canadas-intervention-on-article-36-weapons-reviews/> (accessed November 21, 2016). Weapons reviews would also neglect the security concerns raised by the fully autonomous weapons. In November 2015, India stated, "We feel that LAWS should be assessed not just from the view point of international law including international humanitarian law but also on their impact on international security if there is dissemination of such weapons systems." Statement of India, CCW Meeting of States Parties, Geneva, November 12, 2015, [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/3BAE1E11C60AE555C1257FoF00394109/\\$file/india.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/3BAE1E11C60AE555C1257FoF00394109/$file/india.pdf) (accessed November 27, 2016).

¹²⁹ See, for example, Statements of Canada, Finland, Sweden, and Zambia, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11-15, 2016.

process.”¹³⁰ Such a set of best practices, however, would operate on a voluntary basis and would have less authority than a legally binding instrument.

While strengthening weapons reviews and setting international standards are worthy goals, the CCW meetings about fully autonomous weapons are an inappropriate forum for such discussions. The need to improve reviews is relevant neither specifically nor solely to fully autonomous weapons.¹³¹ Rather, discussions about weapons reviews in the context of fully autonomous weapons distract from the substantive issues presented by the development and use of these weapons.

A binding international ban on fully autonomous weapons would resolve the shortcomings of weapons reviews in this context. A ban would also simplify and standardize weapons reviews by removing any doubts that the use of fully autonomous weapons would violate international law.

Contention #11: Regulation would better address fully autonomous weapons concerns than a ban.

Rebuttal: A binding, absolute ban on fully autonomous weapons would reduce the chance of misuse of the weapons, would be easier to enforce, and would 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 within certain pre-defined parameters.¹³² Such a framework might, for example, limit the use of

¹³⁰ Statement of United States, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument) (November 21, 2016).

¹³¹ Article 36, “Article 36 Reviews,” p. 2.

¹³² 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.”). See also Anderson, Reisner, and Waxman, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*; Gwendelynn Bills, “LAWS unto Themselves: Controlling the Development and Use of Lethal Autonomous Weapons Systems,” *George Washington Law Review*, vol. 83 (2015); Rebecca Crootof, “The Killer Robots are Here: Legal and Policy Implications,” *Cardozo Law Review*, vol. 36 (2015), p. 1879; Elias, “Facing the Brave New World of Killer Robots,” *Indonesian Journal of International & Comparative Law* (2016).

fully autonomous weapons to specific types of locations or purposes. These critics suggest that such an approach would not be over-inclusive because it would more precisely tailor restrictions to the evolving state of fully autonomous weapons technology. Regulations could come in the form of a legally binding instrument or a set of gradually developed, informal standards.¹³³ Whatever its form, however, regulation would not be as effective as a ban.

An absolute, legally binding ban on fully autonomous weapons would provide several distinct advantages over formal or informal regulatory 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 would maximize the stigmatization of fully autonomous weapons, creating a widely recognized norm 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 might be tempted to use the weapons in inappropriate ways in the heat of battle or in dire circumstances (see Contention #2). Furthermore, the existence of fully autonomous weapons would leave the door open 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 the weapons' 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 was 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

¹³³ Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems," *Jean Perkins Task Force on National Security and Law*, p. 22 (explaining that by "'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").

the regulations. Debates about the scope of the regulations and their enforcement would likely ensue.

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.”¹³⁴ 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.¹³⁵ If countries rely on transparency and wait until “norms coalesce” in an admittedly “very difficult area,”¹³⁶ such weapons will likely be developed and deployed, at which point it would probably already be too late to control them.

Contention #12: Ensuring human control during the design and deployment of autonomous weapons would be sufficient to address the concerns they raise.

Rebuttal: In order to avoid the dangers of fully autonomous weapons, humans must exercise meaningful control over the selection and engagement of targets in individual attacks. Only a ban on fully autonomous weapons can effectively guarantee such meaningful control by humans.

Analysis: While there appears to be widespread agreement that all weapons should operate under at least some level of “human control,”¹³⁷ certain critics contend that it need not be directly over individual attacks. These critics argue that human control at the design

¹³⁴ Ibid., p. 20.

¹³⁵ Ibid., p. 25 (referring to the US tendencies toward secrecy).

¹³⁶ Ibid., p. 20.

¹³⁷ At least 26 states described the need to have at least some level of human control or involvement at the 2016 CCW Meeting of Experts on Lethal Autonomous Weapons Systems alone: Austria, Chile, Colombia, Costa Rica, Croatia, Cuba, the Czech Republic, Denmark, Germany, Greece, the Holy See, Ireland, Israel, Italy, Japan, Morocco, the Netherlands, Pakistan, Poland, the Republic of Korea, Sierra Leone, South Africa, Spain, Sweden, Switzerland, and Turkey.

and deployment stages would be sufficient to preempt the concerns associated with fully autonomous weapons because the weapons would operate predictably.¹³⁸ Weapons with such limited control would be unlikely always to operate as expected, however, and human control is not meaningful if there is unpredictability.¹³⁹ Meaningful human control is essential to averting the dangers associated with fully autonomous weapons.

If human control over weapons were confined to the design and deployment stages, unpredictability in weapons would be almost impossible to avoid. Programmers could not always be sure how advanced weapons with complex codes would act in practice. As some scholars note, “[N]o individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways.”¹⁴⁰ In addition, the actions of these weapons could be influenced by factors beyond the programmer. The weapons might rely on dynamic learning processes or processes to adapt existing information for use in new environments.¹⁴¹ The unpredictability of weapons controlled by humans only at the pre-attack stages would indicate that that control was not meaningful.

The absence of meaningful human control would lead to at least three of the fundamental dangers of fully autonomous weapons already outlined in this report. First, because humans could not preprogram fully autonomous weapons to respond predictably to

¹³⁸ United Nations Institute for Disarmament Research, “The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward” (2014), <http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf> (accessed November 21, 2016), p. 3 (noting that some “argue that human control can be sufficiently exercised through the design of a system and by ensuring that it functions reliably and predictably without having a human ‘in the loop’ for each targeting and attack decision.”). See also Adviesraad Internationale Vraagstukken and Commissie Van Advies Inzake Volkenrechtelijke Vraagstukken, “Autonomous Weapon Systems: The Need for Meaningful Human Control,” October 2015, p. 34; Michael C. Horowitz and Paul Scharre, “Meaningful Human Control in Weapon Systems: A Primer,” Center for a New American Security working paper, March 2015, p. 8.

¹³⁹ Even some proponents of fully autonomous weapons concede that the actions of the weapons would be fundamentally unpredictable. See, for example, Bills, “LAWS unto Themselves,” *George Washington Law Review*, p. 197. For a definition of meaningful human control, see Richard Moyes, “Key Elements of Meaningful Human Control,” Article 36 background paper, April 2016, <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> (accessed November 21, 2016), p. 4.

¹⁴⁰ Gary E. Marchant et al., “International Governance of Autonomous Military Robots,” *Columbia Science and Technology Law Review*, vol. 12 (2011), p. 284.

¹⁴¹ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 12. See also Roberto Cordeschi, “Automatic Decision-Making and Reliability in Robotic Systems: Some Implications in the Case of Robot Weapons,” *AI & Society*, vol. 28, no. 4 (2013), p. 436; Gary E. Marchant and Kenneth L. Mossman, *Arbitrary and Capricious: The Precautionary Principle in the European Union Courts* (Washington D.C.: AEI Press, 2004), p. 284.

unforeseeable situations, the weapons would face significant obstacles to complying with international humanitarian or human rights law, which requires the application of human judgment (see Contentions #1 and #5). Second, limiting human control to the design and deployment stages would lead to an accountability gap since programmers and commanders could not predict at those stages how the weapons would act in the field and thus would escape liability in most cases (see Contention #3). Third, fully autonomous weapons would be unable to adhere to preprogrammed ethical frameworks, given their inherent unpredictability,¹⁴² and ceding human control over determinations to use force in specific situations would cross a moral threshold (see Contention #6).¹⁴³

Human control must be exercised over individual attacks in order to be meaningful and address many of the concerns regarding technological advances in weapons systems. Such control would promote legal compliance by facilitating the application of human judgment in specific, unforeseeable situations. It would allow for the imposition of legal liability by creating a link between a human actor and the harm caused by a weapon. Finally, meaningful human control over individual attacks would also ensure that morality could play a role in decisions about the life and death of human beings.

Timeliness and Feasibility of a Ban

Contention #13: It is premature to ban fully autonomous weapons given the possibility of technological advances.

Rebuttal: 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:

¹⁴² See generally Matthias Englert, Sandra Siebert, and Martin Ziegler, “Logical Limitations to Machine Ethics with Consequence to Lethal Autonomous Weapons.”

¹⁴³ Elias, “Facing the Brave New World of Killer Robots,” *Indonesian Journal of International & Comparative Law*, p. 122.

Research 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.¹⁴⁴

This position depends in part on faith that technology could address the legal challenges raised by fully autonomous weapons, which, as explained under Contention #1, seems unlikely and uncertain at best. At the same time, it ignores other dangers associated with these weapons that are not related to technological development, notably the accountability gap, moral objections, and the potential for an arms race (see Contentions #3, 6, and 8).

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 possessed it.

In addition, if ongoing development were permitted, militaries might deploy fully autonomous weapons in complex circumstances with which artificial intelligence could not yet cope. Only after the weapons faced unanticipated situations that they were not programmed to address could the technology be modified to resolve those issues. During that period, the weapon would be likely to mishandle such situations potentially causing great harm to civilians and even friendly forces.

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

¹⁴⁴ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 15.

prevent environmental degradation.”¹⁴⁵ 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 prohibition on a class of weapons. As discussed in Contention #4, in 1995 states parties to the CCW adopted a ban on blinding lasers before the weapons had started to be deployed.¹⁴⁶ During the negotiations, countries expressed many of the same concerns about blinding lasers as they have about fully autonomous weapons, and those negotiations led to a successful new instrument—CCW Protocol IV. States should build on that model and agree to a similar ban on fully autonomous weapons. Although there are differences between the two types of weapons, the revolutionary nature of fully autonomous weapons strengthens, rather than undermines, the case for a preemptive prohibition.¹⁴⁷

Contention #14: A definition of fully autonomous weapons is needed before the concerns they raise can be addressed.

Rebuttal: A common understanding of fully autonomous weapons (also known as lethal autonomous weapons systems) has already largely already been reached, and

¹⁴⁵ Rio Declaration on Environment and Development, U.N. Doc. A/CONF.151/26 (vol. 1), 31 ILM 874, 1992, adopted June 14, 1992, principle 15. The Rio Declaration was a product of the 1992 United Nations Conference on Environment and Development. This 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 November 21, 2016).

¹⁴⁶ CCW Protocol IV.

¹⁴⁷ Human Rights Watch and IHRC, “Precedent for Preemption,” pp. 17-18.

disarmament negotiations have historically agreed on a treaty's detailed, legal definition after resolving other substantive issues.

Analysis: Some critics argue that discussions cannot move toward treaty negotiations without a detailed definition of fully autonomous weapons, also known as lethal autonomous weapons systems (LAWS) by CCW states.¹⁴⁸ For example, one state has noted that “there seemed to be no agreement as to the exact definition of LAWS.... In this regard, many states ... were not supportive of the call made by some states for a preemptive ban on LAWS.”¹⁴⁹ Another has argued that “prohibiting such systems before a broad agreement on a definition would not be pragmatic.”¹⁵⁰ A common understanding, however, should be sufficient to advance deliberations.

Most countries whose statements on the issue are publicly available appear to agree upon the basic elements of what constitutes a fully autonomous weapon. First, they say that fully autonomous weapons, although rapidly developing, remain an emerging technology that does not yet exist.¹⁵¹ Second, they concur that fully autonomous weapons would be, as the name suggests, weaponized or lethal technology.¹⁵²

Third, most of the states that have addressed the topic describe fully autonomous weapons as operating without human control. The terminology employed has varied, from

¹⁴⁸ See, for example, statements made at the 2014, 2015, and 2016 CCW Meetings of Experts on Lethal Autonomous Weapons Systems by Argentina, Australia, China, the Czech Republic, France, Germany, Japan, Morocco, New Zealand, Poland, the Republic of Korea, Spain, Sri Lanka, Sweden, Turkey, and the United Kingdom.

¹⁴⁹ Statement of Israel, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument) (accessed November 21, 2016).

¹⁵⁰ Statement of Turkey, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument) (accessed November 21, 2016).

¹⁵¹ At least 25 states described fully autonomous weapons as emerging or developing technologies during the 2016 CCW Meeting of Experts on Lethal Autonomous Weapons Systems alone. These states include: Algeria, Australia, Austria, Canada, Ecuador, France, Germany, the Holy See, India, Israel, Italy, Japan, Mexico, Morocco, the Netherlands, Norway, Pakistan, Poland, Sierra Leone, South Africa, Spain, Switzerland, Turkey, the United Kingdom, and the United States.

¹⁵² Fully autonomous weapons do not, therefore, encompass surveillance systems or autonomous technology for civilian use. In fact, more than 30 states used the terminology of “lethality” to describe fully autonomous weapons at the 2016 CCW Meeting of Experts on Lethal Autonomous Weapons Systems alone. These states include: Algeria, Austria, Australia, Canada, Costa Rica, Ecuador, Finland, France, Germany, the Holy See, India, Israel, Italy, Japan, Mexico, Morocco, the Netherlands, New Zealand, Norway, Pakistan, Poland, Sierra Leone, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Turkey, the United Kingdom, the United States, and Zambia.

“meaningful human control,”¹⁵³ to “appropriate levels of human judgment,”¹⁵⁴ to “human involvement,”¹⁵⁵ but there seems to be almost universal agreement that fully autonomous weapons lack human control. Finally, while some debate lingers about precisely where human control is absent, agreement is coalescing around the notion that fully autonomous weapons lack human control over the critical combat functions, in particular, over the selection and engagement of targets.¹⁵⁶

Historically in disarmament treaty negotiations, common understandings become detailed legal definitions only at the end of the process. For the Mine Ban Treaty,¹⁵⁷ the Convention on Cluster Munitions,¹⁵⁸ and CCW Protocol IV on Blinding Laser Weapons, the goals, scope, and obligations of the treaty being negotiated were determined before the final definitions. The initial draft text of the Mine Ban Treaty was circulated with definition of antipersonnel landmines from CCW Amended Protocol II.¹⁵⁹ That definition was only a starting point that was revised in later drafts of the text and was still being debated at the final treaty negotiation conference.¹⁶⁰ Similarly, the negotiating history of the Convention on Cluster Munitions began with a declaration in which states at an international conference committed to adopting a prohibition on “cluster munitions that cause unacceptable harm to civilians.”¹⁶¹ While states discussed the definition of cluster munitions at the diplomatic meetings that followed, they did not settle on the definition of cluster munitions to be

¹⁵³ See Contention #12 for further discussion of meaningful human control. See also Human Rights Watch and IHRC, “Killer Robots and the Concept of Meaningful Human Control.”

¹⁵⁴ Statement of the United States, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument) (accessed November 21, 2016).

¹⁵⁵ Statement of Switzerland, CCW Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 11-15, 2016, [http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument) (accessed November 21, 2016).

¹⁵⁶ At least 13 states discussed having control over the critical functions, namely, the selection and engagement of targets, during the 2016 CCW Meeting of Experts on Lethal Autonomous Weapons Systems.

¹⁵⁷ Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction (Mine Ban Treaty), adopted September 18, 1997, entered into force March 1, 1999.

¹⁵⁸ Convention on Cluster Munitions, adopted May 30, 2008, entered into force August 1, 2010.

¹⁵⁹ Stuart Maslen and Peter Herby, “An International Ban on Anti-Personnel Mines: History and Negotiation of the ‘Ottawa Treaty,’” *International Review of the Red Cross*, article no. 325 (1998),

<https://www.icrc.org/eng/resources/documents/article/other/57jpjn.htm> (accessed November 21, 2016).

¹⁶⁰ *Ibid.*

¹⁶¹ Oslo Conference on Cluster Munitions, “Declaration,” February 23, 2007,

<http://www.clusterconvention.org/files/2012/11/Oslo-Declaration-final-23-February-2007.pdf> (accessed November 21, 2016).

adopted until the final negotiations.¹⁶² Working papers and draft protocols from the CCW Group of Governmental Expert meetings about blinding lasers reveal the same pattern: the draft definition contained only the basic elements of the final definition, which would be crafted later in the course of negotiations.¹⁶³

There is already enough international agreement on the core elements of fully autonomous weapons to proceed with negotiations. Getting lost in the details of a definition without first determining the aims of negotiations would be unproductive. It would be more efficient to decide on the prohibitions or restrictions to be imposed on the general category of weapons and then detail to exactly which weapons those prohibitions or restrictions should apply. The international community should, therefore, focus on articulating the goals, scope, and obligations of a future instrument. The final legal definition of fully autonomous weapons can be negotiated at a later stage.

Contention #15: Valuable advances in autonomous technology would be impeded by a ban on the development of fully autonomous weapons.

Rebuttal: A prohibition would not stifle valuable 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.”¹⁶⁴ These critics fear that the ban would therefore impede the exploration of beneficial autonomous technology, such as self-driving cars.

¹⁶² Human Rights Watch, *Meeting the Challenge: Protecting Civilians through the Convention on Cluster Munitions*, November 2010, <https://www.hrw.org/sites/default/files/reports/armsclusters1110webwcover.pdf>, pp. 128-136.

¹⁶³ CCW Group of Governmental Experts, Draft Protocol on Blinding Weapons, CCW/CONF.I/GE/CRP.28, August 12, 1994; CCW Group of Governmental Experts, Various Proposals on Blinding Weapons, CCW/CONF.I/GE/CRP.45, January 11, 1995; CCW Protocol IV, art. 1

¹⁶⁴ Anderson and Waxman, “Law and Ethics for Autonomous Weapon Systems,” *Jean Perkins Task Force on National Security and Law*, p. 14.

In fact, the ban would apply to development only of fully autonomous weapons, that is, machines that could select and fire on targets without meaningful human control. Research and development activities would be banned if they were directed at technology that could be used exclusively for fully autonomous weapons or that was explicitly intended for use in such weapons. A prohibition on the development of fully autonomous weapons would in no way impede development of non-weaponized 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.

Given the importance of keeping fully autonomous weapons out of national arsenals (see Contention #13), a prohibition on development should be adopted, even if it is a narrow one. Including such a prohibition in a ban treaty would legally bind states parties not to contract specifically for the development of fully autonomous weapons or to take steps to convert other autonomous technology into such weapons. It would also create a stronger norm against fully autonomous weapons by stigmatizing development as well as use and could thus influence even states and non-state armed groups that have not joined the treaty.

Contention #16: An international ban on fully autonomous weapons is unrealistic and would be ineffective.

Rebuttal: Past disarmament successes, growing support for a ban, and increasing international discussion of the issue suggest that a ban is both realistic and the only effective 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.”¹⁶⁵ 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).”¹⁶⁶ Other critics believe that

¹⁶⁵ Ibid., p. 3.

¹⁶⁶ 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 November 21, 2016).

even if such a ban could be adopted, it would not be implemented as states would either not join the prohibition or not comply with it.¹⁶⁷ 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, antipersonnel landmines, and cluster munitions, as well as a preemptive ban on blinding lasers, which were still under development. Opponents of the landmine and cluster munitions instruments had frequently said that a ban treaty would never be possible, but the success of these bans has proved their skepticism was misplaced. The number of states joining these treaties and general compliance illustrates the treaties' effectiveness and the ability of humanitarian disarmament to protect civilians from suffering.

Efforts to address the dangers of fully autonomous weapons are following a similar path as previous humanitarian disarmament instruments. 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 International Campaign to Ban Landmines and the Cluster Munition Coalition.

Public support for a ban has bolstered the position of the campaign. As of November 2016, more than 3,000 roboticists and artificial intelligence researchers had signed a 2015 public letter calling for a ban on fully autonomous weapons. According to them, "Just as most chemists and biologists have no interest in building chemical or biological weapons, most AI researchers have no interest in building AI weapons — and do not want others to tarnish their field by doing so, potentially creating a major public backlash against AI that

¹⁶⁷ Anderson, Reisner, and Waxman, "Adapting the Law of Armed Conflict to Autonomous Weapon Systems," *International Law Studies*, p. 398; Allyson Hauptman, "Autonomous Weapons and the Law of Armed Conflict," *Military Law Review*, vol. 218 (2013), p. 195; Aiden Warren and Ingvil Bode, "Altering the Playing Field: The U.S. Redefinition of the Use-of-Force," *Contemporary Security Policy*, vol. 36, no. 2 (2015), p. 191 (arguing that "American decision-makers will not ever consider 'surrendering decision-making regarding the use-of-force and will be even less likely to with the ambiguity deriving from lethal autonomous weapons.'").

curtails its future societal benefits.”¹⁶⁸ Surveys have also revealed support for a ban. For example, a 2015 international survey found that 67 percent of respondents believe that fully autonomous weapons should be internationally banned (see Contention #4).¹⁶⁹ Finally, governments have taken up the debate about fully autonomous weapons. Shortly after civil society pressure began, they added the topic to the CCW agenda, which was significant because the CCW process has previously produced a preemptive ban on blinding lasers and served as an incubator for bans on landmines and cluster munitions. Since 2014, CCW states parties have held three informal experts meetings that have examined the issues surrounding lethal autonomous weapons systems in depth. In the course of these meetings, many states have recognized the need to address these problematic weapons in some way. Fourteen states have expressed explicit support for a ban.¹⁷⁰ States parties that attended the 2016 experts meeting recommended that CCW’s Fifth Review Conference, to be held in December 2016, consider establishing a more formal Group of Governmental Experts to advance discussions.¹⁷¹ Now it is up to the Review Conference to ensure that states pick up the pace and take the next step toward an instrument that bans the development, production, and use of fully autonomous weapons.

Achieving a ban will certainly require significant work and political will. Past precedents and recent developments suggest, however, that a legally binding prohibit on fully autonomous weapons would be the most realistic and effective way to address the dangers these weapons pose.

¹⁶⁸ Future of Life Institute, “Autonomous Weapons: An Open Letter from AI & Robotics Researchers.”

¹⁶⁹ Open Roboethics Initiative, “The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll,” p. 1.

¹⁷⁰ As of November 2016, the states calling for a preemptive ban on fully autonomous weapons were Algeria, Bolivia, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, the Holy See, Mexico, Nicaragua, Pakistan, the State of Palestine, and Zimbabwe. Campaign to Stop Killer Robots, “Ban Support Grows, Process Goes Slowly,” April 15, 2016, <https://www.stopkillerrobots.org/2016/04/thirdmtg/> (accessed November 21, 2016).

¹⁷¹ “Recommendations to the 2016 Review Conference: Submitted by the Chairperson of the 2016 Meeting of Experts,” CCW Meeting of Experts on Lethal Autonomous Weapons Systems, April 2016, [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/6BB8A498BoA12A03C1257FDB00382863/\\$file/Recommendations_LAWS_2016_AdvancedVersion+\(4+paras\)+.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/6BB8A498BoA12A03C1257FDB00382863/$file/Recommendations_LAWS_2016_AdvancedVersion+(4+paras)+.pdf) (accessed November 21, 2016).

Acknowledgments

Bonnie Docherty, senior researcher in the Arms Division of Human Rights Watch and senior clinical instructor at the Harvard Law School International Human Rights Clinic (IHRC), was the lead writer and editor of this report. Joseph Crupi, Anna Khalfaoui, and Lan Mei, students in IHRC, made major contributions to the research, analysis, and writing of the report. Steve Goose, director of the Arms Division, and Mary Wareham, advocacy director of the Arms Division, edited the report. Dinah PoKempner, general counsel, and Tom Porteous, deputy program director, also reviewed the report.

This report was prepared for publication by Marta Kosmyna, associate in the Arms Division, Fitzroy Hepkins, administrative manager, and Jose Martinez, senior coordinator. Russell Christian produced the cartoon for the report cover.

MAKING THE CASE

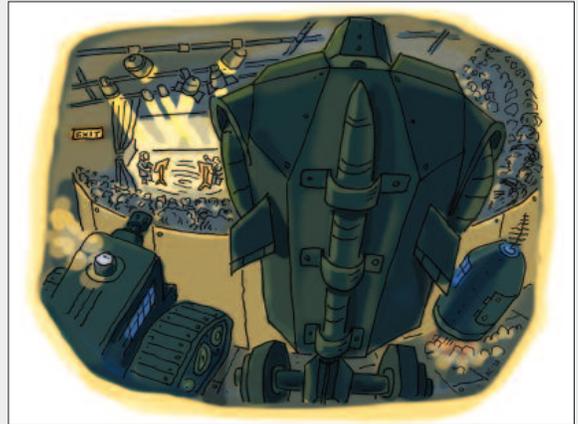
The Dangers of Killer Robots and the Need for a Preemptive Ban

As military technology moves toward ever greater autonomy, experts continue to debate the legality and desirability of fully autonomous weapons. These weapons, also known as “killer robots,” would select and engage targets without meaningful human control. Given that fully autonomous weapons would revolutionize warfare in alarming ways, ongoing discussions are not enough. Preemptive action is of the essence.

Making the Case highlights the dangers of fully autonomous weapons and argues for prohibiting their development, production, and use. In the process, it provides in-depth rebuttals to 16 key critiques of such a ban.

The report details the legal, moral, security, and other threats posed by fully autonomous weapons. The weapons would face significant obstacles to complying with international humanitarian and human rights law and would create a gap in accountability. The delegation of life-and-death decisions to machines would cross a moral threshold. Even the expected military advantages of the weapons could create unjustifiable risks.

Making the Case also examines numerous proposals for dealing with these concerns and finds that only an absolute ban would address them all. The ban should be adopted as soon as possible, before this technology enters military arsenals. Precedent from past disarmament negotiations and treaties shows that a prohibition on fully autonomous weapon would be an achievable and effective solution.



As the debate about “killer robots” continues, the threat they pose looms large.

© 2016 Russell Christian for Human Rights Watch



hrw.org

<http://hrp.law.harvard.edu/clinic>