AUTONOMOUS WEAPONS SYSTEMS AND INTERNATIONAL HUMANITARIAN LAW: 'OUT OF THE LOOP'?

Jarna Petman, LL.D.

ISBN 978-951-51-3963-4 (sid.)
ISBN 978-951-51-3964-1 (PDF)
ISSN 1457-5965

Layout & taitto: Unigrafia Oy
Painatus: Unigrafia Oy
Helsinki 2017

CONTENTS

1 INTRODUCTION

Ours is the age of technology. Today, there are more microprocessors in a regular family car than there were in the Apollo 11 spacecraft that first landed on the Moon in 1969. Indeed, it is not only that technology has become everyday but it is also that everyday has become technology: machines are performing increasingly complex operations on our behalf.[1] Soon, it is predicted, they will be taking care of our elderly; already they are not only driving on our behalf but performing surgical operations without much need for human involvement. With this increasing automatization more and more states have also began to develop weapons systems that would no longer need humans to perform their tasks.[2] Such systems would not be just automated but autonomous.

In a sense, robotic warfare is already reality, for a pilot sitting in an operating room in, say, the United States can already control an unmanned aerial vehicle – a 'drone' – to carry out lethal targeting operations on the other side of the world in, say, Afghanistan. This in fact is the direction to which weapons development has always been moving, with the goal of removing human personnel as far from the risk of harm as possible.[3] The more advanced weapons technology has become, the further away from the battlefield humans have moved. The next step in the 'dehumanization' may remove the human from the process altogether, however. While most military robots, like drones, are currently controlled by a human operator, in the future, it seems likely that some such robots will control themselves.

1 See, e.g., Erik Brynjolfsson & Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (W. W. Norton & Co.: New York, 2014); Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014).

2 See Peter Warren Singer, *Wired For War: The Robotics Revolution and 21st Century Conflict* (Penguin Press: London, 2009).

3 See Peter Asaro, 'How Just Could a Robot War Be?' in Adam Biggle, Katinka Waelbers & Philip A. E. Brey (eds), *Current Issues in Computing and Philosophy* (IOS Press: Amsterdam, 2008) 50–64.

Autonomy can come in varying degrees, though.[4] For example, even simple weapons systems like landmines can be argued to have a very low level of 'autonomy', in the sense that once laid, they are not controlled by the party that laid them but will automatically explode when triggered.[5] And semi-autonomous weapons systems such as fire and forget missiles on aircrafts (that are commonplace in contemporary warfare) lock onto a target identified by the pilot and then attack it without human involvement. Other already existing weapons systems have a higher level of autonomy. For example, both the United States navy and the Israeli army have weapons systems that are able to autonomously detect, track, and fire at incoming missiles, using programmed parameters. Autonomous sentry guns are set up at the border between North and South Korea that will automatically fire at objects within range.[6] But the weapons systems that are currently being designed and developed will have a much higher level of autonomy, with the ability to control their own movement, detect their own targets, and make their own decision to fire at a target and kill, without any human intervention. Weapons systems incorporating strong artificial intelligence may even be capable of learning and making their own decisions about how they decide to conduct targeting.[7]

Such truly autonomous weapons have (possibly) not yet been developed. But already there are drones that are capable of navigating their own flight paths, drones that land and take-off autonomously, and swarming drones that autonomously operate and move together like insects. The prospects that open before us are disconcerting. Autonomous weapons systems raise

4 For a historical overview of the evolution of automation and autonomy in weapons, see Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems', 47 *Vanderbilt Journal of Transnational Law* (2014) 1–54 at 8–10; and for more technical information on autonomous weapons, see Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Chapman & Hall/CRC Press: Boca Raton, FL, 2009) at 7–27.

5 Cf. Paul Scharre & Michael C. Horowitz, *An Introduction to Autonomy in Weapon Systems* (Center for a New American Security Working Paper, Washington, February 2015) at 6–7.

6 Mark Prigg, 'Who goes there? Samsung unveils robot sentry that can kill from two miles away'. *Daily Mail*, 15 September 15 2014 available at <www.dailymail.co.uk/sciencetech/article-2756847/Who-goes-Samsung-reveals-robot-sentry-set-eye-North-Korea.html>.

7 See, e.g., Robert Sparrow, 'Killer Robots', 24 *Journal of Applied Philosophy* (2007) 62–77 at 64–65.

questions that are by nature not only technical or military, but also ethical, socio-political and legal. Governments, experts and non-governmental organizations alike are deeply divided as to whether the development of autonomous lethal systems is useful, legal, ethical or even desirable.

On the one side, there are those that emphasize the benefits of autonomous weapons systems.[8] Autonomous systems can detect and process complex information in enormous speed; they make both decision-making and strikes more flexible, faster and more precise; as they replace human fighters, they save lives by reducing the number of casualties; unlike humans, robots will not necessarily be driven by a need to protect themselves, meaning that they have more scope to act conservatively and in a self-sacrificial manner in cases where target identification is uncertain or where acting in self-defence would result in excessive civilian harm; as they do not have feelings, they can without worry be made to take on tasks that are dull, dirty and dangerous; and as they do not have emotions to cloud their judgment, such as fear, anger, selfishness, revenge, hysteria or frustration, and as they do not feel hunger or fatigue, they may lead to end-results that are less harmful than those resulting from human action. There would presumably be no senseless atrocities like the My Lai massacre or Fallujah in an autonomous robot war.

8 See, e.g., Michael N. Schmitt, 'Autonomous Weapons Systems and International Humanitarian Law: A Reply to Critics', *Harvard National Security Journal* (2013) 1–37; Kenneth Anderson and Matthew C. Waxman, 'Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation Under International Law', in Roger Brownsword, Eloise Scotford & Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017) 1097; Jakob Kellenberger, 'International Humanitarian Law and New Weapon Technologies' (Keynote Address by the President of the International Committee of the Red Cross, 34th Round Table on current issues of international humanitarian law, San Remo, 8–10 September 2011), 94 *International Review of the Red Cross* No. 886 (2012) 809–813.

On the other side of the divide, there are those that see autonomous weapons systems as a threat.[9] Human life has no value when a machine makes the decision to kill; like any technology, autonomous weapons will lead to arms race and, like any technology, autonomous weapons systems will be vulnerable to abuse and misuse; such systems may be faulty and thus come with severe malfunctions, leading to serious failures; in relieving those troops that are in possession of autonomous weapons systems from immediate risk whilst putting their enemy fighters and civilians into real danger, autonomous weapons turn battles disproportionate and unfair; unlike humans, robots do not have such positive feelings as compassion or pity, they show no mercy; this may even apply to those persons that originally programme the parameters for a machine, for they are so far removed from the battles and the targets – both physically and emotionally – that they may be wholly indifferent towards them. Battles become unreal; enemies become inhuman, irrelevant.

These are the two sides of the argument that is currently being waged on the international sphere as experts, governments and non-governmental organizations are trying to position themselves with regard to autonomous weapons systems. In terms of international law, this debate about how to deal with weapons that are capable of making decisions has over the last few years focused more on the sphere of international humanitarian law (*jus in bello*) than that of the law governing the use of force between states (*jus ad bellum*). That this should be so, owes to the fact that the rules on the use of force between states apply to every use of force in international

9 See, e.g., Noel E. Sharkey, 'The evitability of autonomous robot warfare', 94 *International Review of the Red Cross* No. 886 (2012) 787–799; David Akerson, 'The Illegality of Offensive Lethal Autonomy' in Dan Saxon (ed.) *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff/ Brill: Leiden & Boston, 2013) 65–98; Jeroen van den Boogaard, 'Proportionality and Autonomous Weapons Systems', 6 *Journal of International Humanitarian Legal Studies* (2015) 247–283; Kjølv Egeland, 'Lethal Autonomous Weapon Systems under International Humanitarian Law¨', 85 *Nordic Journal of International Law* (2016) 89–119; Daniele Amoroso & Guglielmo Tamburrini, 'The Ethical and Legal Case Against Autonomy in Weapons Systems', 17 *Global Jurist* (2017) issue 3; see also Future of Life Institute, 'An Open Letter to the United Nations Convention on Certain Conventional Weapons', 21 August 2017, available at <futureoflife.org/autonomous-weapons-open-letter-2017>; and Samuel Gibbs, 'Elon Musk leads 116 experts calling for outright ban of killer robots', *The Guardian*, 20 August 2017, available at <www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war>.

relations, regardless of the type of weapons being deployed. They therefore also apply to weapon systems incorporating a greater or lesser degree of autonomy.

In short, the international legal framework for the use of force provides that any use of force in international relations is prohibited. The prohibition, enshrined in Article 2(4) of the United Nations Charter, is universally accepted as a norm of customary international law,[10] and is often considered a rule of peremptory international law that is to trump every other conflicting rule and from which there is to be no derogation (except in strictly limited extraordinary circumstances).[11] The 'force' in the prohibition is understood to mean armed force and thereby exclude purely economic, diplomatic or political coercion.[12] Armed force can take the form of incursions of military forces into another state's territory, or even cross-border shooting into foreign territory, regardless of whether armed confrontation results, and even if troops withdraw immediately.[13] In addition, the prohibition includes indirect force,[14] typically manifested by a state's participation in organized armed groups' or allied states' use of force on another state's territory,

10 The International Court of Justice (ICJ) has on several occasions confirmed the customary law character of the prohibition, see, e.g., *Military and Paramilitary Activities in and against Nicaragua (Nicaragua* v. *United States) (Jurisdiction and Admissibility)*, ICJ Reports (1984) 39 at para. 73, and *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territories (Advisory Opinion)* ICJ Reports (2004) 136 at para. 87; see also Oliver Dörr, 'Use of Force, Prohibition', *Max Planck Encyclopedia of Public International Law*, September 2015, *Oxford Public International Law Online* (Oxford University Press 2015).

11 As much was confirmed by the ICJ in *Military and Paramilitary Activities in and against Nicaragua (Nicaragua* v. *United States) (Merits)*, ICJ Reports (1986) 14 at 100; see also Lauri Hannikainen, *Peremptory Norms* (Jus Cogens) *in International Law: Historical Development, Criteria, Present Status* (Lakimiesliiton kustannus: Helsinki 1988); and Tarcisio Gazzini, *The Changing Rules on the Use of Force in International Law* (Manchester University Press 2005) at, e.g., 2, 25, 89, and 122.

12 Cf. Olivier Corten, 'The Controversies over the Customary Prohibition on the Use of Force: A Methodological Debate', 16 *European Journal of International Law* (2005) 803–822.

13 See Albrecht Randelzhofer & Oliver Dörr 'Purposes and Principles, Article 2(4)' in Bruno Simma et al. (eds) *United Nations Charter: A Commentary* (3rd ed., Oxford University Press 2012), vol. I, at 216. See also Tom Ruys, 'The Meaning of "Force" and the Boundaries of the *Jus Ad Bellum*', 108 *American Journal of International Law* (April 2014) 159–210.

14 See Declaration on Principles of International Law Concerning Friendly Relations and Cooperation Among States in Accordance with the Charter of the United Nations, GA Res. 2625 (XXV), 2 October 1970, 121 (Annex), UN Doc. A/8028 (1971).

such as arming and training armed groups that actually use or threaten force against that state.[15] Allowing one's territory to be used to commit acts of force against another state, for example by hosting the infrastructure for launching autonomous weapons, could also conceivably constitute an indirect use of force.[16] It has been suggested, however, that 'non-lethal activities' such as 'leadership training, organizational assistance, political or economic intelligence gathering, political subversion, or information operations' do not cross the use-of-force threshold, while the provision of 'lethal ("military") training and logistical support, such as instruction on the use of weapons or transporting of rebel forces during operations, would, by contrast, be an unlawful use of force'.[17] The provision of targeting intelligence would appear to fall under the latter category. While the prohibition also covers *threats* of force, these have rarely led any state to argue that the prohibition has been violated.[18]

When a state consents to another state's use of force on its territory, the prohibition is not violated.[19] Commonly, consent will take the form of a request that another state's armed forces intervene in an internal armed conflict or fight against alleged terrorists. According to the International Law Commission, there can be no ambiguity about the existence of consent, which 'must be freely given and clearly established. It must be actually expressed by the State rather than merely presumed on the basis that the State would have consented if it had been asked'.[20] In cases in which a state

15 See *Nicaragua Case (Merits)*, *supra* note 10 at para. 228.

16 See Randelzhofer & Dörr 'Purposes and Principles', *supra* note 12 at 211.

17 Michael N. Schmitt and Andru E. Wall, 'The International Law of Unconventional Statecraft', 5 *Harvard National Security Journal* (2014) at 375.

18 Randelzhofer & Dörr 'Purposes and Principles', *supra* note 12 at 218.

19 As has been noted by the International Law Commission, '[v]alid consent by a State to the commission of a given act by another State precludes the wrongfulness of that act in relation to the former State to the extent that the act remains within the limits of that consent'; such consent to resort to force can only be granted by the highest government authorities and must be given freely, clearly, and in advance or at the time of the operation, see *Commentary to International Law Commission Draft Articles on State Responsibility*, art. 20. See also *Nicaragua Case (Merits)*, *supra* note 10 at para. 246. A state cannot however consent to violations of international human rights or humanitarian law being committed on its territory; see Heyns, *Report of the Special Rapporteur*, *infra* note 65 at para. 84.

20 See *Commentary to ILC Draft Articles on State Responsibility*, *supra* note 18.

has not given its consent to the use of force on its territory, the UN Charter foresees two exceptions that are applicable today to the prohibition of the use of force, namely Security Council action or a Security Council mandate authorising the use of force to maintain or restore international peace and security pursuant to Chapter VII of the UN Charter; or individual or collective self-defence of one or more states against an armed attack.[21] Both two exceptions are subject to further conditions. The use of force on the basis of a Security Council mandate must be consistent with the conditions and objectives of the mandate in question, while the right of self-defence can only be invoked in the case of an actual or imminent armed attack. The principles of necessity, proportionality and immediacy apply to every use of force between states.[22] In the case of self-defence, the force must be necessary to repel the actual or imminent attack as well as proportionate.[23]

When one of the internationally recognised legal bases for the use of force is invoked, the legality of that use of force depends on the particular circumstances of the case. In this, the nature of the weapon system being deployed is irrelevant. In situations where the Security Council has issued a mandate to take all necessary measures to counter 'a threat to the peace, a breach of the peace or an act of aggression', the deployment of any weapon system (autonomous or otherwise), will be examined for compatibility with the mandate. The decision to deploy a particular weapon system will influence the legality of the use of force in exceptional cases only. For

21 Cf. Anders Henriksen, '*Jus ad bellum* and American Targeted Use of Force to Fight Terrorism Around the World', 19 *Journal of Conflict & Security Law* (2014) 211–250. At times two additional but controversial exceptions have also been invoked: humanitarian intervention to prevent or put an end to massive violations of human rights and, within narrow conditions, limited force to protect or rescue a state's own nationals on the territory of another state; see Randelzhofer & Dörr 'Purposes and Principles', *supra* note 12 at 222–28.

22 See *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, ICJ Reports (1996), 226 at para. 41; see also *Nicaragua Case (Merits)*, *supra* note 10 at para. 194; *Case Concerning Oil Platforms (Islamic Republic of Iran* v. *United States of America)*, ICJ Reports (2003) 161 at paras 43, 51, 73–77; and *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo* v. *Uganda)*, ICJ Reports (2005) 168 at para. 147. See also Yoram Dinstein, *War, Aggression, and Self-Defence* (Cambridge University Press, 2012) at 233, 267.

23 See Albrecht Randelzhofer & Georg Nolte, 'Ch.VII Action with Respect to Threats to the Peace, Breaches of the Peace, and Acts of Aggression, Article 51' in Bruno Simma et al. (eds) *United Nations Charter: A Commentary* (3rd ed., Oxford University Press 2012), vol. II, 1397.

example, deploying nuclear weapons in response to a less than massive conventional attack would almost certainly be regarded as disproportionate and therefore unlawful.[24] In addition, certain weapons fall under an arms control regime that totally bans specific weapons (such as, say, chemical and biological weapons) or subjects them to specific restrictions. In such cases, possession or deployment may constitute a violation of the regime in question. As a rule, however, the legality of deploying a specific weapon is not regulated by *jus ad bellum*.

Nevertheless, autonomous weapons systems do pose a direct and distinct challenge to the *jus ad bellum* proportionality principle. Proportionality is closely linked to other *jus ad bellum* principles and it requires the careful analysis of the overall consequences of the proposed use of autonomous weapons systems. If the principle of proportionality cannot be satisfied, then the other principles governing a lawful use of force cannot be met. A state in possession of an autonomous weapons arsenal will have the advantage of using such weapons in defence, in particular in cases where the other side lacks the same level of technology. As autonomous weapons save soldiers' lives and this would weigh heavily in favour of deploying such weapons, for any harm and injury thus caused would be blamed on the unlawful aggressor and not on the defending state. The same applies to potential collateral damage, as such damage may be justified as unintentional and in pursuance of the legitimate military objective of self-defence. The ability to use autonomous weapons against an unlawful threat must therefore be seen as a benefit in the proportionality calculation of a state.

In all, the presence of autonomous weapons systems may influence the choice of a nation to use force in two ways: they not only directly threaten the sovereignty of potentially hostile nations, but they may make it easier for the decision-makers to opt for the deployment of armed force. The availability of autonomous weapons is feared to considerably lower the threshold for initiating an armed conflict as the troops of states that

24 Cf. Advisory Council on International Affairs & Advisory Committee on Issues of Public International Law, *Autonomous Weapons Systems: The Need for Meaningful Human Control* (No. 97 AIV / Np. 26 CAVV, October 2015, The Hague) at 18–19.

deployed such weapons systems would be at lower immediate risk[25] – the increasing deployment of armed drones in remote conflict areas is considered an early example of this phenomenon.[26] Moreover, as the future autonomous weapons are likely to be capable of learning and adapting their functioning in response to changing circumstances in the environment in which they are deployed, as well as making firing decisions on their own, such weapons could be directly responsible for starting an armed conflict independent from – and contrary to – the political and strategic convictions of the state in possession of them.[27]

* * *

As states now consider how to deal with the problems posed by autonomous technologies, the debates have been particularly heated concerning the challenge that autonomous technologies pose to international humanitarian law. No wonder, for autonomous weapons systems present a unique regulatory problem for laws of war because they replace the human role in war and killing. Indeed, the role autonomous weapons systems play as a *weapon* may not be as disconcerting as the role that they play as their own *operator*.

This research report will explore the legal implications of autonomous weapons systems for international humanitarian law. Throughout, the analysis will be confined to *lethal* autonomous weapons systems. While non-lethal robots raise their own concerns, they will remain outside the scope of the report. The focus will be on the legal implications of transferring the decision to kill from human to autonomous machines, rather than on the more general implications of the automatization of technology. For this

25 See, e.g., Pablo Kalmanovitz, 'Judgment, Liability and the Risks of Riskless Warfare' in Nehal Bhuta et al. eds, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 145–163; and Nathalie Weizmann & Milena Costas Trascasas, *Autonomous Weapon Systems under International Law* (Academy Briefing No. 8, Geneva Academy of International Humanitarian Law and Human Rights, 2014) 9.

26 See, e.g., U. C. Jha, Winf Commander (retd), *Killer Robots: Lethal Autonomous Weapon Systems' Legal, Ethical and Moral Challenges* (Vij Books India Pvt Ltd: New Delhi, 2016) at 71.

27 See, e.g., Heather M. Roff, 'Lethal Autonomous Weapons and *Jus Ad Bellum* Proportionality', 47 *Case Western Reserve Journal of International Law* (2015) 37–52.

reason, also any analysis of military technology that continues to include a human actor in the decision-making loop, such as drones, will remain outside the scope of the research. The starting assumption is that it is only with the complete removal of the human actor that the fundamental bases for the principles of international humanitarian law are called into question.

2 AUTONOMOUS WEAPONS SYSTEMS DEFINED

A state's choice of methods and means of warfare is a limited one.[28] A weapon may be problematic under or contrary to international humanitarian law because of the manner in which it is used. More rarely, the weapon itself may be inherently problematic. There are, at present, three reasons for banning a weapon under international humanitarian law.[29] Weapons that are incapable of distinguishing between military targets, on the one hand, and civilians and civilian objects, on the other, are prohibited; such weapons include bacteriological weapons that will inevitably spread and infect also the civilian population, and certain types of mines and booby traps. Weapons that cause unnecessary suffering or excessive injury to enemy combatants are prohibited; bullets that explode on contact with the human body, for example, and laser weapons that cause permanent blindness. Weapons the effects of which cannot be controlled and which thus result in indiscriminate harm to soldiers and civilians alike are prohibited; a computer virus that is deployed to knock out an opponent's military communication system is likely, fundamentally uncontainable as it is, to also knock out, say, the communication system of the emergency services. In view of this, international agreements exist to specifically regulate a number of problematic weapons, such as expanding bullets,[30]

28 See Protocol additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (opened for signature 8 June 1977, entered into force 7 December 1978) 1125 *United Nations Treaty Series* 3 (Protocol I) Art 35(1); see also Convention (IV) respecting the Laws and Customs of War on Land and its annex on Regulations concerning the Laws and Customs of War on Land (The Hague, 18 October 1907) Art 22.

29 The ensuing examples are from AIV & CAVV, *Autonomous Weapons Systems*, *supra* note 23 at 19–20.

30 Declaration (IV,3) concerning Expanding Bullets, The Hague, 29 July 1899.

poisonous gases,[31] landmines,[32] and blinding lasers.[33] Also, to keep track of advances in weapons technology, international humanitarian law requires states parties to the Geneva Conventions to ensure in the study, development, acquisition, or adoption of a new weapon, that it would not in any circumstances be prohibited by international law.[34]

The legal requirement to conduct a review of new weaponry is understood to rely on the assumption that the existing principles of international humanitarian law apply *ipso facto* also to new weaponry and technological developments in warfare. And yet, regardless of the universal consensus that the law of armed conflict continues to apply to new developments, the method of merely subsuming new technologies under pre-existing rules raises questions as to whether this is sufficient in terms of legal clarity, and whether this suffices in view of the specific characteristics and the humanitarian impact that the new technology may have. These very questions have gained urgency with the introduction of 'autonomous weapons systems' into the battlesphere, for such systems will profoundly influence the nature of warfare and, potentially, the nature of international legal regulation. For this, the future challenges raised by the development of autonomous weapons systems are now at the forefront of international legal discourse.

The concept of 'autonomous weapons systems' is understood to refer to robotic weapons that, once activated, can select and engage targets without further intervention by a human operator. Such systems are equipped with sensors that enable them to have a degree of situational awareness, with computers that process the information gathered from the surroundings,

---

31 Protocol for the Prohibition of the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, Geneva, 17 June 1925.

32 Convention on the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction (opened for signature 18 September 1997, entered into force 1 March 1999) 2056 *United Nations Treaty Series* 211; Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on 3 May 1996 annexed to the Convention on Certain Conventional Weapons (*infra* note 133), Geneva, 3 May 1996, entry into force 3 December 1998, 2048 *United Nations Treaty Series* 93.

33 Protocol to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects (Protocol IV, entitled Protocol on Blinding Laser Weapons), 13 October 1995, entered into force 30 July 1998, 33 *International Legal Materials* 1218.

34 Article 36 of Additional Protocol I to the Geneva Conventions, *supra* note 27.

and with 'effectors' (weapons) that implement the 'decisions' taken by the computers. While there is, as of yet, no internationally agreed definition of an 'autonomous weapon', all the current definitions adopted by governments, experts and non-governmental organizations alike capture this characteristic nature of autonomous weapons systems: humans are no longer required in the targeting decision-making process.[35]

The current development of autonomous weapons systems (AWS) has the potential of moving in two different directions: either as the extension of human soldiers, or as the replacement of humans in the battlefield by autonomous agents.[36] The choice, in other words, is between 'weapons that augment our soldiers and those that can become soldiers'.[37] At present, the prevailing view seems to be that robots will be used only to augment and extend soldier's involvement in war. As long as AWS continue to be regarded as tools to further distance humans from combat, they can be considered but the latest technological advancement in the long process of weapons development rather than a radically new category of weaponry in need of novel legal responses. Nevertheless, the idea that autonomous weapons systems will come to replace soldiers continues to gain traction. In this view, the very point of AWS is for them to be more than a mere extension of humans as they will by themselves have the potential to make the decision to kill without human involvement. And herein lies the crux of the matter: while, for example, the current use of drones may be criticized for various other reasons,[38] their *capability* of adhering to the principles of international humanitarian law is uncontroversial because humans are

35 See, e.g., United States Department of Defense, 'Autonomy in Weapon Systems', Directive Number 3000.09, 21 Nov 2012, <www.dtic.mil/whs/directives/corres/pdf/300009p.pdf>; HRW & IHRC, *Losing Humanity*, *infra* note 63 at 2; see also United Kingdom Ministry of Defence, 'Unmanned Aircraft Systems', Joint Doctrine Publication 0–30.2, 12 September 2017, at 13 and 43, para. 4.17, available at <www.gov.uk/government/uploads/system/uploads/attachment_data/file/640299/20170706_JDP_0-30.2_final_CM_web.pdf>.

36 Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate Publishing: Surrey, 2009) at 35.

37 Major David F Bigelow, 'Fast forward to the robot dilemma', *Armed Forces Journal* (1 November 2007) available at <armedforcesjournal.com/fast-forward-to-the-robot-dilemma>.

38 See, e.g., Chris Downes, '"Targeted Killings" in an Age of Terror: The Legality of the Yemen Strike', 9 *Journal of Conflict and Security Law* (2004) 277–294.

involved in the targeting process – a fully autonomous weapons system, in contrast, would take human operators out of the decision-making loop altogether, thus marking a fundamental shift in the age-old goal of weapons development of distancing humans from war through technology. [39] What, if any, implications should such a fundamental shift have on the legal regulation of warfare?

Much of the current legal discourse on autonomous weapons systems is forward-looking in the sense that fully autonomous intelligent systems are still (it is assumed) only being developed for military purposes in the future. The incremental development of AWS is expected to, first, enhance aspects of operations such as take off and navigation, leading to full autonomy over time.[40] As technological advances are made, more and more sophisticated sensing and computational systems will be implemented. While AWS are not yet a reality on the battlefield, the level of autonomy in weapons has been growing steadily.[41] Indeed, while some experts confine themselves to the view that the introduction of full autonomy to weapons systems is inevitable and imminent,[42] others emphasize that the technology required for 'fully autonomous military strikes' has existed for years already.[43] Given that 'the military application of autonomous systems is inevitable and already pursued by some nations', NATO launched in the summer of 2013 an un-classified Multinational Capability Development Campaign among nineteen nations (one of the contributing nations being Finland), NATO and the European Union so as to 'improve awareness and understanding of autonomous systems, promote interoperability and provide guidance for the development of, use of, and defence against, au-

39 See Markus Wagner, Taking Humans Out of the Loop: Implications for International Humanitarian Law', 21 *Journal of Law, Information and Science* (2011) 155 at 157–158.

40 Noel Sharkey, 'Automating Warfare: Lessons Learned from the Drones', 21 *Journal of Law, Information and Science* (2011) 140 at 141.

41 See, e.g., Timothy Coughlin 'The Future of Robotic Weaponry and the Law of Armed Conflict: Irreconcilable Differences?', 17 *UCL Jurisprudence Review* (2011) 67–99.

42 See, e.g., Gary E. Marchant et al., 'International Governance of Autonomous Military Robots', 12 *Columbia Science and Technology Law Review* (2011) 272.

43 See the commentary by the former chief scientist of the US Air Force Werner J.A. Dahm, 'Killer Drones Are Science Fiction', *The Wall Street Journal,* 15 February 2012, at 11.

tonomous systems'.[44] In a similar vein, the US Undersecretary of Defence for Acquisition, Technology and Logistics announced in the fall of 2014 the commissioning of a new study focusing on 'the science, engineering, and policy problems that must be solved to permit greater operational use of autonomy across all warfighting domains';[45] the results of the study were published in June 2016.[46]

To be sure, semi-autonomous systems are already widely present in the battlefield, providing intelligence gathering, surveillance and reconnaissance, as well as target acquisition, designation and engagement capabilities.[47] Limited autonomy is also already present in fire-and-forget munitions, loitering torpedoes, and intelligent antisubmarine or anti-tank mines, to name a few examples, as well as in systems, such as the Phalanx Close-in Weapons Systems used in the *Aegis* class cruisers by the United States Navy, that are capable of autonomously performing their own 'search, detect, evaluation, track, engage and kill assessment functions'.[48] While South Korea has developed robotic sentries with 'automatic surveillance' to monitor the demilitarized zone with North Korea, Russia has been developing ground sentry robots to guard missile sites, albeit that their degree of autonomy remains unclear.[49] The United Kingdom, for its part, has been developing a new semi-autonomous aircraft, *Taranis*, that has been described as 'an

44 See *Proceedings Report of the Multinational Capability Development Campaign (MCDC) 2013–2014: Focus Area 'Role of Autonomous Systems in Gaining Operational Access'* (MCDC Secretariat/NATO Allied Command Transformation, December 2014) at 4 and 6, available at <innovationhub-act.org/sites/default/files/AxS_Product_2light.pdf >.

45 Frank Kendall, Under Secretary of Defense for Acquisition, Technology and Logistics, 'Terms of Reference – Defense Science Board 2015 Summer Study on Autonomy', Memorandum for Chairman, Defense Science Board, 17 November 2014, available at <dspace.mit.edu/bitstream/handle/1721.1/86935/Advisory%20_%20Defense%20Science%20Board%20_%20Memo%20on%20AI%20Roadmap.pdf?sequence=166>.

46 Defense Science Board, *Report of the Defense Science Board 2015 Summer Study on Autonomy* (Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, Washington, DC, 2016) available at <www.hsdl.org/?view&did=794641>.

47 For selected examples of human-supervised autonomous weapons systems, see Scharre & Horowitz, *An Introduction to Autonomy*, *supra* note 5 at 21–23.

48 See United States Navy, 'Phalanx Close-in Weapons Systems', *United States Navy Fact File*, <www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2>.

49 David Hambling, 'Armed Russian robocops to defend missile bases', *New Scientist*, 23 April 2014, <www.newscientist.com/article/mg22229664-400-armed-russian-robocops-to-defend-missile-bases>

autonomous and stealthy unmanned aircraft'.[50] Although humans will remain in the loop for the time being, it may well be capable of autonomous flight. An operational derivative of *Taranis'* platform is expected to enter into military service after 2030.[51] The United States too has been developing a semi-autonomous drone, the *X-47B*, which is to be able to take off and land without human input. Although the current development of *X-47B* does not reportedly envision autonomous target selection, it will be capable of semi-autonomous flight.[52] The development of full autonomy has however been included in all the roadmaps of the United States forces since 2004,[53] and the US Air Force *Unmanned Aircraft Systems Flight Plan 2009–2047* suggests that fully autonomous flight systems will be possible by 2025.[54] Indeed, as noted by a US Air Force officer in a study in 2005: '[it] is not a matter of "will" we employ [autonomous weapons]; it is a matter of "when" we employ them'.[55] That such anticipation is shared by a great number of states, is evident in, for example, in the 2011 study by Noel Sharkey, computer scientist and the chair of the International Committee for Robots Arms Control, in which he analysed official robotics development reports revealing that over fifty countries were at the time engaged in developing autonomous weapons systems.[56]

Continued advances in autonomy have already resulted in changes involving military tactics. Moreover, they have already underlined the po-

50 BAE Systems, 'Taranis', <www.baesystems.com/product/BAES_020273/taranis>.

51 See 'Video: UK ends silence on Taranis testing', *FlightGlobal: Pioneering Aviation Insight*, 5 February 2014, <www.flightglobal.com/news/articles/video-uk-ends-silence-on-taranis-testing-395574>.

52 See Northrop Grunman, 'Unmanned Combat Air System Carrier Demonstration', <www.northropgrumman.com/Capabilities/X47BUCAS/Documents/X-47B_Navy_UCAS_FactSheet.pdf>; cf. also James Holmes, 'The Mighty X-47B: Is It Really Time for Retirement?', *National Interest*, 25 May 2015.

53 See Sharkey, 'Automation and Proliferation', *infra* note 50, at 231.

54 United States Air Force, *Unmanned Aircraft Systems Flight Plan 2009–2047*, 18 May 2009, Washington D.C., available at <www.fas.org/irp/program/collect/uas_2009.pdf>, at 50.

55 Major Michael A. Guetlin, *Lethal Autonomous Weapons – Ethical and Doctrinal Implication*s, JMO Department, Naval War College, 2005, available at <www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA464896>, at 18

56 Noel Sharkey, 'The Automation and Proliferation of Military Drones and the Protection of Civilians', 3 *Law, Information and Technology* (2011) 229 at 235; cf. also Scharre & Horowitz, *An Introduction to Autonomy*, *supra* note 5 at 12.

tential benefits of autonomous systems to the military. A member of the now-defunct US Joint Forces Command summed up the benefits of AWS in 2005 as follows: 'They don't get hungry. They're not afraid. They don't forget orders. They don't care if the guy next to them has just been shot. Will they do a better job than humans? Yes.'[57] AWS are also considered cheaper to operate than human-operated weapons, and are capable of operating continuously, without the need for rest.[58] Moreover, fewer humans are needed for the operation of AWS.[59] In a similar fashion as it is already possible for a single operator to manage a swarm of semi-autonomous drones, it will be possible for a single human commander to assign mission parameters to AWS and monitor them from a safe distance. In this regard, one of the major weaknesses of the current remotely piloted vehicles is the ease with which the enemy may interfere with their satellite or radio links; AWS are seen to alleviate this concern, as they will be capable of operating without continuous contact with home base.[60] Also, while remotely piloted systems currently have a delay time of approximately 1.5 seconds, limiting their effectiveness in a higher tempo battle space, AWS are potentially capable of processing battlefield information considerably faster and more efficiently than humans.[61]

Autonomous weapons systems are thus expected to allow for multiplication of force, expansion of the battlespace, extension of the fighter's reach, the ability to respond faster to an ever-increasing battlefield tempo, and greater precision due to 'persistent stare' (i.e. constant video surveillance that allows more time for decision making and more eyes on target).[62] As they can be fitted with any variety of sensory technologies including

57 See Tim Weiner, 'New Model Army Soldier Rolls Closer to Battle', 16 February 2005, available at <www.nytimes.com/2005/02/16/technology/new-model-army-soldierrolls-closer-to-battle.html>.

58 Guetlin, *Lethal Autonomous Weapons*, *supra* note 49, at 2.

59 ISee, e.g., Marchant et al., 'International Governance', *supra* note 41 at 275.

60 See Noel Sharkey, 'Saying No! to Lethal Autonomous Targeting', 9 *Journal of Military Ethics* (2010) 369 at 377.

61 See, e.g., Guetlin, *Lethal Autonomous Weapons*, *supra* note 49, at 4–5.

62 See, e.g., Ronald Arkin, 'Lethal Autonomous Systems and the Plight of the Non-combatant', *AISB Quarterly* [Society for the Study of Artificial Intelligence and the Simulation of Behaviour] (2013) 1–9, at 1.

infrared vision, sonar, high definition cameras, and sophisticated auditory sensors, this will give AWS an advantage over human sensory capabilities. Because robotic sensors will be better equipped to make battlefield observations than humans, proponents of AWS suggest that autonomous weapons may in fact be more capable of adhering to the principles of international humanitarian law than human soldiers.[63] In this regard it is also considered important that, unlike humans, robots will not necessarily be driven by a need to protect themselves, meaning that they have more scope to act conservatively and in a self-sacrificial manner in cases where target identification is uncertain or where acting in self-defense would result in excessive civilian harm.[64] While this seems a logical argument, it may be questionable how realistic it really is in light of the cost of such systems and the fact that one of the major drivers for increased autonomy is the perceived need for robots to defend themselves when they lose contact with a human operator.[65] Another perceived advantage of AWS is their lack of emotion. Unlike humans, they can be designed without emotions to cloud their judgment, such as fear, anger, hysteria, or frustration.[66] They will thus be also immune to the psychological problem of 'scenario fulfilment', that is to say, to the tendency to believe in something and consider only information favourable to this and ignore/deny information against it.[67] This, it has been suggested, would also lead to reduction in friendly casualties.[68] The verity of such claims has yet to be demonstrated. (I will return to them shortly, in the chapters to follow.)

From the perspectives of military and economic efficiency alike, it seems inevitable that lethal autonomous weapons systems will be developed and deployed – unless limited by international humanitarian law. Indeed, as

63 See Thomas K. Adams, 'Future Warfare and the Decline of Human Decision Making', 31 *Parameters: US Army War College Quarterly* (2001–2002) 57–71 at 57–58.

64 See, e.g., Marchant et al., 'International Governance', *supra* note 41 at 279–280.

65 See Singer, *Wired for War*, *supra* note 2, at 127; Peter Warren Singer, 'The Ethics of Killer Applications: Why is It so Hard to Talk about Morality when it comes to New Military Technology', 9 *Journal of Military Ethics* (2010) 299, at 303–304

66 See, e.g., Singer, *Wired for War*, *supra* note 2, at 63; Marchant et al., 'International Governance', *supra* note 41 at 280;

67 Cf. *infra* note 118 and the accompanying text.

68 See, e.g., Arkin, 'Lethal Autonomous Systems', *supra* note 56, at 1.

emphasized by the United States Air Force *Unmanned Aircraft Systems Flight Plan 2009–2047*, if and when the increased autonomy of weapons systems is deemed necessary to meet 'extreme performance parameters' and the full autonomy thereof is the long-term goal, then the legal and ethical implications of these developments will have to grappled with: 'Authorizing a machine to make lethal combat decisions is contingent upon political and military leaders resolving legal and ethical questions.'[69] As human involvement diminishes, the difficulties faced by autonomous weapons systems in adhering to the principles of international humanitarian law will become more and more significant, requiring a thorough legal analysis.

69 US Air Force, *Flight Plan 2009–2047*, *supra* note 48, at 30.

3 THE COMPLIANCE OF AUTONOMOUS SYSTEMS WITH THE LAW OF ARMED CONFLICT

The international community is currently divided as to how and to what effect the principles of international humanitarian law may be applied on autonomous systems. For example, in November 2012, Human Rights Watch together with the International Human Rights Clinic of the Harvard Law School issued a report, *Losing Humanity: The Case against Killer Robots*, that called for a complete ban on the production and use of autonomous weapons systems, for they 'would not only be unable to meet legal standards but would also undermine essential non-legal safeguards for civilians'.[70] A few days later, the United States Department of Defence released a directive outlining the Department's policies on the development and use of such autonomous systems, setting out a review and regulatory process within the existing legal framework to ensure that whatever level of autonomy a weapon system might have, the autonomous function has been subjected to legal review to ensure that it complies with the laws of war.[71] In April 2013, the United Nations special rapporteur on extrajudicial, summary or arbitrary executions called for a moratorium on the development of autonomous weapons systems until a legal framework is developed.[72] While the three actors disagreed on the solution, they all started from the assumption that autonomous weapons systems will raise challenges of adherence to international humanitarian law.

As noted, a weapon may present problems for or be contrary to international humanitarian law because the weapon itself may be inherently problematic. Consequently, Article 36 of Additional Protocol I to the Geneva Conventions requires states parties to ensure in the 'study, development, acquisition, or adoption of a new weapon' that it would not, 'in some or all circumstances', be prohibited by the Protocol or any other rule of in-

70 Human Rights Watch [HRW] & International Human Rights Clinic [IHRC], *Losing Humanity: The Case Against Killer Robots*, available at <www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots> at 1–2.

71 US DoD, 'Autonomy in Weapon Systems', *supra* note 34.

72 Christof Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, UN Human Rights Council, UN Doc. A/HRC/23/47, 9 April 2013.

ternational law.[73] While a number of states, including some not party to Additional Protocol I, have adopted such mechanisms in principle,[74] Article 36 has not thus far been particularly effective in addressing concerns about autonomous weapons.[75] Indeed, if the contrary were true, this would be a cause for astonishment, for very few states are currently actively engaged with weapons review processes; where such reviews have been carried out in practice, they have not resulted in a prohibition of development, acquisition, nor adoption of any new weapon.[76] As consequence, Article 36 can hardly be argued to be a customary norm of international law. In view of this, the International Committee of the Red Cross (ICRC) has made the more general – and more plausible – argument that the systematic assessment of the legality of all new weapons is an obligation on all states based on the prohibition on illegal weapons and the restrictions on means and methods of warfare.[77]

Given the non-specific formulation of Article 36 and the lack of state practice from which to draw, however, the exact scope of the obligation and the practical requirements as to how comply with it remain unclear. What is clear, is that the provision does not concern only inherently unlawful weapons but also covers those that in principle have the capacity for great precision but may in practice also be used in ways that are inconsistent with international humanitarian law – autonomous weapons are likely to

73 See Protocol I additional to the Geneva Conventions, *supra* note 27, Art. 36, the full text of which reads: '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.'

74 See ICRC, *ICRC Study on Customary International Humanitarian Law*, Rule 71, available at <www.icrc.org/customary-ihl/eng/docs/v1_cha_chapter20_rule71>.

75 See the discussion in Chantal Grut, 'The Challenge of Autonomous Lethal Robotics to International Humanitarian Law', 18 *Journal of Conflict & Security Law* (2013) 5–23 at 20–22.

76 As noted by Chantal Grut, blinding lasers are the only weapons system that was regulated at the international level before being significantly deployed to ill effect, see *ibid.* at 9 fn. 28.

77 See International Committee of the Red Cross, 'A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977' 88 *International Review of the Red Cross* (2006) 933.

belong to this latter category. In the matter of automation in particular, the ICRC commentary to Protocol I notes as follows:

> *The use of long distance, remote control weapons, or weapons connected to sensors positioned in the field, leads to the automation of the battlefield in which the solider plays an increasingly less important role. The counter-measures developed as a result of this evolution, in particular electronic jamming (or interference), exacerbates the indiscriminate character of combat. In short, all predictions agree that if man does not master technology, but allows it to master him, he will be destroyed by technology.*[78]

According to the ICRC, any legal review of a new weapon should also include consideration of the technical performance of the weapon as this is relevant in determining whether its use may cause indiscriminate effects.[79]

However, the unique regulatory problems that truly autonomous weapons systems present arise not so much from their nature as weapons, but from their replacement of the human role in war and killing: an autonomous system is capable of operating itself, selecting and engaging targets without further intervention by a human operator – potentially independently from any human oversight. The crux of the problem is seen to lie in that, once humans have been taken out of the command loop, autonomous weapons systems 'appear to be incapable of abiding by key principles of international humanitarian law'.[80] This is because dehumanized autonomous systems – regardless of the level and sophistication of their artificial intelligence – are considered incapable of ever adequately making the kinds of highly complex and contextual analyses that international humanitarian law requires, such as being able to distinguish civilian objects from legitimate military targets. Taking humans 'out of the loop' has not only raised questions of the compatibility of autonomous weapons systems with the fundamental

78 ICRC, *Commentary to the Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, at para 1476.

79 ICRC, 'A Guide to the Legal Review', *supra* note 70, at para. 1.3.2.

80 HRW & IHRC, *Losing Humanity*, *supra* note 63 at 30.

requirements of international humanitarian law but has also complicated allocation of responsibility for war crimes and crimes against humanity. And yet, some advocates have suggested that autonomous systems might already be able to perform in compliance with international humanitarian law better than humans can, for example in the speed with which they can respond to threats.[81] Increasingly, military systems are becoming 'too fast, too small, too numerous and will create environments too complex for humans to direct'.[82]

What kind of implications, then, might increasing weapons autonomy have for the law of armed conflict? In essence, the bedrock rules of international humanitarian law that apply to every combat operation are those of distinction (only legitimate targets may be attacked), proportionality (any incidental or collateral damage inflicted on civilians who are not directly participating in the hostilities must not be excessive in relation to the military advantage thus obtained), and precaution (during the planning and execution of an attack, precautionary measures must be taken at all times to minimise – and if possible to nullify – the effects of the attack on civilians and civilian objects).

Compliance with the Principle of Distinction?

There are two components to the principle of distinction (sometimes referred to as discrimination). First, parties to armed conflict must be able to distinguish between civilians and enemy combatants. Second, parties to armed conflict must be able to distinguish between civilian and military objects. Only military objectives may be targeted in an attack.[83] This principle is codified in Article 48 of the Additional Protocol I to the Geneva Conventions. For states that are not signatories to the Additional Protocol

81 See Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 7–8 (and references in notes 14–20).

82 Adams, 'Future Warfare', *supra* note 63 at 58.

83 Protocol I additional to the Geneva Conventions, *supra* note 27, Arts 48 and 51, especially 51(4). See also *ICRC Study on Customary International Humanitarian Law*, *supra* note 67, Rule 75.

I, the principle applies as customary international law.[84] According to the ICRC, Article 48 of Additional Protocol I reflects the foundational principle of the laws and customs of war that civilians must be protected and therefore must be distinguished from combatants.[85] In addition, the International Court of Justice has held that the rule against indiscriminate attacks is a 'cardinal' principle of international humanitarian law.[86] Weapons which are by nature indiscriminate – whether in all or only some circumstances – are prohibited, in both international[87] and non-international[88] armed conflicts. While adherence to the principle of distinction has become increasingly difficult as the nature of military operations has evolved from state-against-state warfare to counterinsurgency operations, such challenges have not, however, changed the core of the principle: parties to a conflict must distinguish between civilian targets and military targets.[89]

On its face, the principle of distinction seems like a simple enough black-and-white rule: either a potential target is or is not a military target. Difficulties arise however from the fact that a target can be classified as both a civilian and military target depending on the context. Indeed, the kind of analysis that is generally required by the principle of distinction is highly complex and highly contextual, that is to say, the kind of analysis that the human mind is uniquely adept at. It is debatable whether AWS will ever have the same level of ability to distinguish civilian objects from legitimate

84 See, e.g., Gary D. Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (Cambridge University Press, 2012) at 2.

85 ICRC, *Commentary to the Protocol*, *supra* note 71, at 598.

86 See *Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion)*, ICJ Reports (1996) at 78.

87 Protocol I, *supra* note 27, Art 51(4); Rome Statute of the International Criminal Court (17 July 1998, into force 1 July 2002) 2187 *United Nations Treaty Series* 90, Art 8(2)(b)(xx); see also *ICRC Study*, *ibid.*, Rule 71.

88 *Ibid.* Cf., Jimena M. Conde Jiminián, 'The Principle of Distinction in Virtual War: Restraints and Precautionary Measures under International Humanitarian Law' (2010–2011) 15 *Tilburg Law Review* 69–91, 75.

89 See, e.g., Solis, *The Law of Armed Conflict*, *supra* note 77, at 254.

military targets.[90] This is particularly the case in many of the highly asymmetric conflicts which are prevalent today, in which it can be extremely difficult to distinguish a farmer digging a trench from a member of an armed group planting an improvised explosive device; and with regard to which disagreement continues to persist as to the exact circumstances under which it is legal to lethally target civilians involved to varying degrees in an armed conflict.

Specifically, three main concerns regarding the ability of AWS to distinguish legal targets from civilian targets have been identified.[91] First, AWS may be susceptible to 'weak machine perception'. Second, AWS may have difficulties in interacting with their environment, leading to the 'frame problem'. Third, there may be a problem of 'weak software'.

Weak Machine Perception

Distinction requires an evaluation that is based on sensory input. While existing technology is now approaching the ability to distinguish between a human and a non-human object, it is still far from being able to make the necessary distinction between a civilian and a combatant with sufficient clarity. Already this purely material capability of differentiation is a technological challenge that some robotics experts believe to be insurmountable.[92] But even presupposing major advances in sensor technolo-

90 Richard M. O'Meara, 'Contemporary Governance Architecture Regarding Robotics Technologies: An Assessment' in Patrick Lin, Keith Abney and George A. Bakey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass., 2012) 129 at 165; J. McClelland 'The Review of Weapons in Accordance with Article 3 of Additional Protocol I', 85 *International Review of the Red Cross* (2003) 397 at 408; and ICRC, *International Humanitarian Law and the Challenges of Contemporary Armed Conflicts* (ICRC Report of the 31st International Conference of the Red Cross and Red Crescent, 28 November – 1 December 2011) at 40. For the argument that robotic systems will never have the kind of capabilities needed, see Noel E. Sharkey, 'Cassandra or False Prophet of Doom: AI Robots and War', 23 *IEEE Intelligent Systems* (2008) 14 at 16–17 and his 'The Ethical Frontiers of Robotics', 32 *Science* (2008) 1800 at 1800–01.

91 In this, I follow the threefold distinction made by Armin Krishnan in his *Killer Robots*, *supra* note 35, at 35.

92 See, e.g., Sharkey, 'Automating Warfare', *supra* note 39, at 143–44; and Sharkey, 'The evitability of autonomous robot warfare', *supra* note 9, at 788.

gy, the question remains whether the highly complex appraisal processes and value judgments that the duty of distinction requires could ever be handled by algorithms.

The problem is compounded in non-international armed conflicts that have now come to be the typical situation of armed conflict. In them, targets are only lawful if they are directly engaging in hostile activity or have the intention of engaging in hostile activity. Notions such as 'asymmetrical warfare' and 'urban warfare' are indicative of the difficulties that may be involved in differentiating between irregular fighters and other legitimate targets, on the one hand, and the civilian population, on the other. As non-uniformed combatants in an armed conflict are identified by their engagement or intention to engage in hostilities,[93] distinction cannot be a mere matter of sensors identifying particular weaponry or enemy uniforms but becomes a matter of interpreting human behaviour. A targeting decision by AWS would thus have to be based on situational awareness and an understanding of human intention.[94] It is unclear whether artificial intelligence will ever be capable of complying with the imperative of distinction in such situations of profound confusion. Discussing such situations, some commentators have pointed to the criterion of 'gut feeling' that is expressly included in, for example, the US army guidelines concerning ethical conduct on combat missions, dictating that as the final mental step before deciding whether to shoot, the soldier is to decide whether a particular action is morally 'right'.[95] Even advocates of AWS acknowledge that such deliberation may not be amenable to algorithmic programming.[96]

It seems particularly problematic to program rules of behaviour to ambivalent combat situation. One solution to this problem would be to only allow AWS to fire when they have been fired upon. Another would be to limit the use of AWS to situations where the declared hostile force is easily recognizable.[97] Yet another suggested solution is to have the AWS target

93 See Solis, *The Law of Armed Conflict*, *supra* note 77, at 254–255.

94 See Sharkey, 'Saying No!', *supra* note 54, at 379.

95 Geiß, *The International-Law Dimension of AWS*, *supra* note 135 at 14.

96 Arkin, *Governing Lethal Behavior*, *infra* note 192 at 51.

97 Major Jeffrey S Thurnher, 'No One at the Controls: Legal Implications of Fully Autonomous Targeting', 67 *Joint Forces Quarterly* (2012) 77 at 83.

the weapon, rather than the human holding the weapon: any injury to the human would technically be collateral damage.[98] Although theoretically possible, this solution does not adequately address the possibility of indiscriminate attacks. Distinguishing between a weapon and any other object may be just as difficult as distinguishing between a civilian and an enemy combatant.[99] Think, for example, of a scenario in which enemy combatants force civilians into transporting weapons for them: although under international humanitarian law they would not be legal targets, under the latter proposal they would be treated as collateral damage.

Yet, the demand that robots target only other weapons systems, not humans, is arguably one of the reasons why existing autonomous systems like the Phalanx have not been regarded so concerning.[100] Phalanx is designed only to shoot down missiles heading towards a ship, and the civilian objects that could, on the ocean, be mistaken for an incoming missile are limited (but do occur, as the case of the *USS Vincennes* illustrates, see below). To be sure, if for example an autonomous drone is only programmed to fire at tanks, then the scope for mistakenly targeting civilian objects is significantly reduced (albeit not eliminated, for it may be impossible for a robot to tell apart for example a tank that was a museum exhibit, or a tank that had been abandoned in a civilian area). The range of circumstances in which targeting only weapons systems could sufficiently avoid the risk of mistaken targeting may, however, be limited. It is one thing to use an automatic anti-missile defence system in the middle of an ocean, it would be another thing entirely to target weapons systems in built up civilian areas, or to target weapons which might not necessarily be military objects, such as civilian guns. In this respect, the risks of undermining the principle of distinction may be more serious in the case of AWS that are mobile and increasingly capable of, say, directing their own flight paths, as compared

98 Chief Engineer John S. Canning, 'A Concept of Operations for Armed Autonomous Systems: The difference between "Winning the War" and "Winning the Peace"', Naval Surface Warfare Center/ Defence Technical Information Center presentation 2006, available at <www.dtic.mil/ndia/2006disruptive_tech/canning.pdf>

99 Krishnan, *Killer Robots*, *supra* note 35, at 106.

100 Cf. Canning, 'A Concept of Operations', *supra* note 89; Wagner, 'Taking Humans Out of the Loop', *supra* note 38, at fn 29.

to, for example, automatic sentry guns which could be placed only in well-marked areas where the presence of civilians is highly unlikely.[101]

Frame Problem

The frame problem refers to the challenge of how to limit the scope of the reasoning that is required to derive the consequences of a particular action. That is to say, how to represent the effects of actions in algorithm so as to enable AWS to make decisions on the basis only of what is relevant to an ongoing situation without having explicitly to consider all that is not relevant. In complex and fast-paced modern battlefields, AWS will have difficulty in assessing all the information needed to correctly interpret the situation. Running through all the possible scenarios is not a viable option for this, in theory, would take an infinite time. Consequently, AWS will have to be programmed to distinguish between relevant and irrelevant information. In an open, equivocal environment, however, programming this type of distinction could lead to situations where the information is incorrectly interpreted which in turn might lead to an indiscriminate attack. Because of the frame problem, AWS could either be too slow to be militarily effective in an actual combat mission, or be prone to indiscriminate action because of missing important details or incorrectly interpreting the changing situations.[102]

The frame problem is complicated by the presumption codified in Article 50(1) of Additional Protocol I, dictating that an attack will be unlawful where there is lack of certainty as to the legitimacy of the target: 'In case of doubt whether a person is a civilian, that person shall be considered to be a civilian'.[103] The mere existence of some doubt does not bring the presumption into operation; instead, the doubt must cause 'a reasonable

101 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 12.

102 Krishnan, *Killer Robots*, *supra* note 35, at 99.

103 Additional Protocol I to the Geneva Conventions, *supra* note 28, Art. 50(1): 'A civilian is any person who does not belong to one of the categories of persons referred to in Article 4 A (1), (2), (3) and (6) of the Third Convention and in Article 43 of this Protocol. In case of doubt whether a person is a civilian, that person shall be considered to be a civilian'.

attacker in the same or similar circumstances to hesitate before attacking'.[104] The threshold is framed in terms of human reasonableness, which quite complicates its adoption with regard to autonomous weapons systems. This determination is highly contextual and would require different 'doubt thresholds' depending on the circumstances. It has been suggested that as long as human operators do not program the 'doubt thresholds' unreasonably high (thereby making it more likely for the AWS to attack), autonomous weapons systems will not violate the principles of distinction.[105] A suggestion has also been made that the principle 'first, do no harm' could be programmed into an autonomous system, which would always prohibit it from automatically deploying deadly force if the situation cannot be classified unequivocally.[106]

Weak Software

The more complicated software becomes, the less predictable it will be. As no single programmer understands or knows the entire piece of software, interactions within it become unpredictable too.[107] Combined with the open environment of an asymmetric conflict, this could lead to situations where AWS apply force indiscriminately because of an unanticipated software error.

The potential 'weak software' problems faced by programming AWS may be exemplified by the difficulties faced by the programmers of the chess-playing computer *Deep Blue* which was created in 1997 and which would eventually beat the world-class chess player Garry Kasparov. As the computer gradually became a more capable chess player than the programmers themselves were, it became increasingly difficult for them to tell whether a particular move by *Deep Blue* was a bug or good tactics.[108] The

104 Schmitt, 'Autonomous Weapons Systems', *supra* note 7, at 16.

105 Schmitt, 'Autonomous Weapons Systems', *supra* note 7, at 17.

106 Arkin, *Governing Lethal Behavior*, *infra* note 194 at 58.

107 Krishnan, *Killer Robots*, *supra* note 35, at 100.

108 Nate Silver, *The Signal and the Noise: Why Most Predictions Fail but Some Don't* (Penguin Press: New York, 2012) at 286.

same difficulty will be faced when programming AWS: at a certain point, it will become difficult to tell whether an autonomous weapons system is making an error, or whether it is seeing and reacting to something that is beyond immediate human grasp.

* * *

Although the principle of distinction appears to be an objective requirement, it has subjective elements that create challenges for the use of autonomous weapons systems. The gravest concern is that even the most intelligent IT programs and sensors will not be able to provide the autonomous systems with the ability of understanding the context, of interpreting (human) intentions and emotions, and of identifying (in situations of asymmetric conflict), who is and who is not a combatant – thereby rendering the autonomous systems unable to apply the status of non-combatant in practice.[109] Accordingly, the sensory technology must first develop sufficiently so that AWS have enough information to be capable of reliably distinguishing between civilian and military targets, and that information must then be processed in an efficiently and accurately enough manner, so that mistakes are not made and the AWS do not target in an indiscriminate fashion. These steps however require technology that does not yet (seem to) exist. While it is conceivable that the technology required to make the necessary distinction will eventually be developed, the fundamental challenges that inhere in the removal of humans from the targeting decision cannot be ignored simply because the technology may exist someday.

It should be noted that the above issues are somewhat unique in the context of developing AWS. Weapons are usually criticized as indiscriminate because of a lack of sophistication and precision, such as has been the case with landmines and cluster bombs, for example. Autonomous weapons systems, on the other hand, are being designed as extremely sophisticated

109 See, e.g., Peter Asaro 'On banning Autonomous Weapons Systems: Human rights, automation and the dehumanisation of lethal decision-making', 94 *International Review of the Red Cross* No. 886 (2012), 687–709, Yoram Dinstein, 'The Principle of Distinction and Cyber War in International Armed Conflicts', 17 *Journal of Conflict and Security Law* (2012) 261–277.

systems that do have the capacity of undertaking distinction – the problem is that they may simply not do the job as well as human-controlled weapons can. It is precisely in this regard that the role autonomous weapons systems play as a weapon is not as concerning as the role that they play as their own operator.

Compliance with the Principle of Proportionality?

While the principles of international humanitarian law strive to protect civilian populations, there is no way to eliminate civilian death and injury from war altogether. The principle of proportionality seeks to address the protection of civilians directly, mandating that where collateral damage to civilians occurs, it must be proportional to military advantage: an attack is thus prohibited if the incidental civilian harm is excessive in relation to the concrete and direct military advantage anticipated by the attack.[110] Parties to armed conflict are required to take all feasible precautions to ensure that targets are in fact military objectives, to avoid civilian harm, and to ensure compliance with the principle of proportionality.[111]

Although there is no reference to proportionality in Additional Protocol II, which applies to non-international armed conflicts,[112] the ICRC has argued that because proportionality in inherent to the principle of humanity – which is included in the Protocol's preamble – it must be included *ipso facto* in the Protocol's application.[113] Moreover, as the ICRC could find no official practice contrary to the principle of proportionality in either international or non-international armed conflicts, the principle has in its view crystallized into customary law.[114] As for the notion of 'military advantage', it has been interpreted by many states to include both the

110 Protocol I, *supra* note 27, Art 51(5)(b); *ICRC Study*, *supra* note 67, Rule 14.

111 Protocol I, *supra* note 27, Art 57; *ICRC Study*, *supra* note 67, Chapter 5.

112 Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), 8 June 1977, entered into force 7 December 1978, 1125 *United Nations Treaty Series* 609.

113 See *ICRC Study*, *supra* note 67, at 48.

114 See *ICRC Study*, *supra* note 67, at 48.

particular advantage anticipated from an attack, as well as the advantage anticipated to the military operation as a whole.[115] The military advantage must be 'concrete and direct'. According to the ICRC the 'concrete and direct' indicates that the advantage 'must be substantial and relatively close, and that advantages which are hardly perceptible and those which would only appear in the long term should be disregarded'.[116]

Proportionality requires a contextual weighing of two factors: the possibility of harm to civilians and civilian objects, on the one hand, and the potential military advantage of the attack, on the other. The determination of the potential harm to civilians is more readily capable of an objective determination. Consequently, collateral damage simulators are already used by commanders to ensure that attacks are proportional.[117] The determination of military advantage, however, is more contextual and discretionary. In determining whether the military advantage requirement has been met, one asks if a 'reasonable commander' would have come to a similar conclusion. In this vein, the Trial Chamber of the International Criminal Tribunal for the former Yugoslavia described criminal accountability for a disproportionate attack as depending on the inherently human 'reasonable person' standard: 'whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack'.[118] The evaluation of military advantage on the basis of the reasonable commander allows for operational discretion.

Adherence to the principle of proportionality requires a subjective assessment. The practical application thereof requires a weighing of potentially competing interests: military advantage and the protection of civilians. This weighing of interests is only possible on a case-by-case basis: different circumstances require different responses. It is this contextual and discre-

115 See *ICRC Study*, *supra* note 67, at 49.

116 ICRC, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Yves Sandoz, Christophe Swinarski, & Bruno Zimmermann eds, Martinus Nijhoff: Geneva, 1987) at 598.

117 Schmitt, 'Autonomous Weapons Systems', *supra* note 7, at 19–20.

118 *Prosecutor* v *Galic (Judgment)* ICTY-98-29 (5 December 2003), 43 *International Legal Materials* 794 at 808.

tionary nature of proportionality that leads to concerns that AWS may be incapable of adhering to the principle. In fact, the question of whether an autonomous system could undertake a proportionality analysis may be even more fraught than the one concerning the ability of AWS to comply with the principle of distinction. To start off with, there is no clear metric as to what the principle of proportionality might require. The proportionality analysis is too highly contextual to allow for it to be reduced, for example, to a rule that you can have one civilian casualty per one combatant killed; or two civilian casualties per a unit commander; or three civilian casualties per one tank destroyed. Instead, it requires an assessment of the level of civilian harm, an assessment of the value of the military advantage gained, and additionally requires consideration of whether there are any alternative avenues to gaining that military advantage which would result in less civilian harm.

In addition to the highly contextual, and inevitably somewhat subjective question of the 'reasonableness' of the decision-maker, another factor in the equation is the extent to which armed forces (or AWS themselves) are required to put themselves at risk in order to avert civilian casualties on the other side; this being an element of the requirement to take precautions in attacks and to 'do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects'.[119] The extent to which armed forces (or their AWS) are required to take risks upon themselves in order to avoid civilian casualties is a truly complex question on which there is no agreement.[120]

According to some commentators, the proportionality rule has no direct application to weapons development, due to the requirement for complex, value-based, case-by-case determination in which the circumstances will have to be weighed in their totality.[121] However, unlike other weapons, AWS could replace the human decision maker. In questioning the legality

119 Protocol I, *supra* note 27, Art 48; and the discussion above.

120 Cf. *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia* (2000) 39 *International Legal Materials* 1257, at 1271.

121 William Boothby, *Weapons and the Law of Armed Conflict* (Oxford University Press, 2009) at 7.

of AWS, proportionality must therefore be considered. While humans are known to be capable of balancing complex interests, the same cannot – for the time being – be said of AWS. As the evaluation of proportionality requires relative weight be placed on competing interests, AWS should be able to anticipate the effect of all potential decisions and the potentially resulting number of civilian casualties. They would also have to react to changing circumstances, whereupon they would have to be able to calculate the military advantage, and determine whether the collateral damage is acceptable. [122]

Systems that determine the likelihood of collateral damage already exist and are used to determine what level of command is required to authorize an attack. A commander then weighs the potential collateral damage against military advantage.[123] The 'frame problem' applies in this context: to calculate the collateral damage potentially ensuing from an attack, AWS will either have to calculate the consequences of every possible action (and thus taking an infinite amount of time), or make assumptions that could potentially lead to a disproportionate attack. To be sure, determinations of collateral damage will always involve assumptions, for certainty is almost never possible in armed conflict. In the event that AWS are employed in open civilian environments, due care must be taken to ensure that information on which to base the assumptions required to determine collateral damage is sufficient and adequately collected and processed.

Currently, there are no systems that would be capable of calculating military advantage. Proponents of AWS suggest, however, that they are possible in theory.[124] Again, the 'frame problem' complicates any determination of military advantage, for the decision-maker has to consider both the immediate and the long-term consequences of the planned action. While human decision-makers are capable of such consideration, it has yet to be demonstrated that this can be replicated by software. Since AWS will not have an infinite amount of time to make these calculations, some shortcuts and preconceived understandings will have to be programmed

122 Wagner, 'Taking Humans Out of the Loop', *supra* note 38, at 159, 162.

123 Schmitt, 'Autonomous Weapons Systems', *supra* note 7, at 19–20.

124 Schmitt, 'Autonomous Weapons Systems', *supra* note 7, at 19–20.

into the software. These shortcuts may lead to errors, however, and thus to disproportionate attacks.

* * *

In an armed conflict, anticipated military advantage and possible collateral damage are constantly shifting and entirely dependent on context, thereby requiring complex processing and sensing capabilities, as well as an algorithm that is capable of making speedy and correct determinations of proportionality. Consequently, in considering the compliance of AWS with the principle of proportionality, the concern is that although it is possible for autonomous systems to be programmed to observe the principle of proportionality in many respects, and to minimize collateral damage by selecting appropriate weapons or munitions and properly directing them, it may not be possible for the systems to make the qualitative and subjective decision that damage to civilians exceeds or outstrips the military advantage provided by the attack.[125] Given that the principle of proportionality cannot be reduced to any kind of clear-cut formula,[126] and that there is no agreement on the articulation of the edges of the prohibition of an excessive attack, it is difficult to imagine how one could even begin to try and write the prohibition into a piece of software coding.[127] Even if a rule could be formulated that was capable of being written into code, one would then come back to some of the same difficulties faced in the context of distinction analysis: it is unclear how a AWS could go about accurately determining which of the objects in its range were civilian, which were military, and which might be dual use, so that it could even go on to weigh the proportionality question. With the current level of technological development, one cannot say that AWS will never be able to

125 See, e.g., Sharkey, 'The evitability of autonomous robot warfare', *supra* note 8, at 789; Patrick Lin, George Bekey & Keith Abney, 'Robots in War: Issues of Risk and Ethics' in R. Capurro & M. Nagenborg (eds) *Ethics and Robotics* (AKA Verlag: Heidelberg, 2009) 49–67, at 57–58.

126 Solis, *The Law of Armed Conflict*, *supra* note 77, at 274.

127 Wagner, 'Taking Humans Out of the Loop', *supra* note 38, at 8–9; Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 13.

make these kinds of determinations satisfactorily, albeit that the obstacles may presently seem unconquerable.

Compliance with the Principle of Precaution?

The principle of precautions in attack was first set out in the 1907 Hague Convention (IX) concerning Bombardment by Naval Forces in Time of War.[128] The demand that parties to an international armed conflict take certain precautions whenever they initiate an attack was also confirmed in Article 57 of Additional Protocol I that requires that '[i]n the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects'.[129] This requirement that is closely linked to the imperative of distinction and the principle of proportionality is now 'most probably' considered customary international law.[130] And yet, the obligation imposed on belligerent parties is 'essentially relative in nature, as situations may arise when civilians simply cannot be spared'.[131]

The active precautions that parties must take when planning or executing an attack have been specified in the Additional Protocol I as follows: do everything feasible to verify that objectives constitute military objectives; take all feasible precautions in choosing the means and methods of attack so as to avoid or minimize collateral damage to civilians and civilian objects; and refrain from launching attacks expected to be in breach of the principle of proportionality.[132] While 'feasible' precautions may be difficult to define with any precisions, they have been described in the 1980 Convention on Conventional Weapons as 'those precautions which are practicable or practically possible taking into account all circumstances ruling

128 Convention (IX) concerning Bombardment by Naval Forces in Time of War (The Hague, 18 October 1907, entry into force on 26 January 1910), Art. 2(3).

129 Additional Protocol I to the Geneva Conventions, *supra* note 28, Art. 57(1).

130 Gary D. Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (2nd ed., Cambridge University Press, 2016) at 521.

131 Frits Kalshoven, *Reflections on the Law of War: Collected Essays* (Martinus Nijhiff/Brill: Leiden, 2007) at 546; see also Yoram Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (2nd ed., Cambridge University Press 2010) at 138.

132 Additional Protocol I to the Geneva Conventions, *supra* note 28, Art. 57(2)(a).

at the time, including humanitarian and military considerations'.[133] As for non-international armed conflicts, in addition to an overall demand in Additional Protocol II that 'the civilian population and individual civilians shall enjoy general protections against the dangers arising from military operations',[134] there are now more specific requirements of precaution in attack included in the more recent treaty law, namely in the Amended Protocol II to the Convention on Conventional Weapons and in the Second Protocol to the Hague Convention for the Protection of Cultural Property.[135] Given that the requirement is contained also in other instruments pertaining to non-international armed conflicts as well as in military manuals which are applicable or have been applied in non-international armed conflicts, it is now considered to be a customary norm on a par with the duty to take precautions in attack in international armed conflicts.[136]

The requirement to take precautions applies to the whole planning phase of an armed deployment and concerns all persons involved in preparations, that is to say, not only commanders but also, arguably, the manufacturers and programmers of autonomous weapons systems.[137] The requirement of care goes even further, however, as the original planning must also be 'valid and decisive after the mission has begun'.[138] Accordingly, given that various unforeseen situations and challenges may occur in the course of a combat mission, some authors argue that the very principle of precaution implicitly gives rise to the duty to keep a human soldier always at least 'on the loop' – with supervisory control, that is – so as to enable her to res-

133 United Nations Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, adopted 10 April 1981, New York, entered into force 2 December 1983, 19 *United Nations Treaty Series* 1823 (1990), Art. 3(4).

134 Additional Protocol I to the Geneva Conventions, *supra* note 108, Art. 13(1).

135 Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices *supra* note 32, Art. 3(10); and Second Protocol to The Hague Convention of 1954 for the Protection of Cultural Property in the Event of Armed Conflict, The Hague, 26 March 1999, entry into force 9 March 2004, UNESCO Doc. HC/1999/7, Art. 7.

136 *ICRC Study on Customary International Humanitarian Law*, *supra* note 74, Rule 15.

137 See William H. Boothby, *Conflict Law: The Influence of New Weapons Technology* (T.M.C. Asser Press: Den Haag, 2014) at 115.

138 Robin Geiß, *The International-Law Dimension of Autonomous Weapons Systems* (Friedrich Ebert Stiftung Study, Berlin, October 2015) at 16, available at <www.fes.de/ipa>.

pond to the various new situations as they develop.[139] Those that are of a different opinion point to the fact that computer-guided weapons systems process information and react to their surroundings much more rapidly than human beings ever can, which in their view renders it highly questionable whether any soldier remaining 'on the loop' in a combat situation would be capable of intervening early enough where AWS was preparing to violate a rule of international humanitarian law.[140]

To be sure, the duty of precaution implicates measures that are highly context-dependent and susceptible to rapid and unpredictable change. They include the continuous verification of the target, the choice of weapon and the timing and method of the attack. Attacks must be suspended or cancelled if it becomes clear that they will have a disproportionate impact or if the target is not (or no longer) legitimate. Also, where possible, civilian populations must be issued a warning prior to an attack. Because of the contextual nature of the requirement of precaution, and the continued re-assessment that it implies, many commentators have come to the conclusion that at least for the next ten years, autonomous weapons systems will not in the vast majority of cases be able to perform the required assessment independently, without human intervention.[141] It is because of this need for continuous re-assessment that those weapons that currently exist with autonomous functions are primarily deployed against military platforms such as military aircraft, warships and military vehicles in environments or situations in which there is little to no risk of collateral damage to civilians or civilian objects or inaccurate assessments. While the deployment of defensive weapons with autonomous functions against enemy projectiles or missiles raises few if any problems under international humanitarian law,

139 Milena Costas Trascasas & Nathalie Weizmann, *Autonomous Weapon Systems under International Law* (Academy Briefing No. 8, Weizmann Geneva Academy of International Humanitarian Law and Human Rights, November 2014, at 4, available at <www.geneva-academy.ch/our-projects/publications>.

140 Philip Alston, 'Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law', 21 *Journal of Law, Information & Science* (2012) 35–60 at 54.

141 AIV & CAVV, *Autonomous Weapons Systems*, *supra* note 24 at 26; see also Jonathan David Herbach, 'Into the Caves of Steel: Precaution, Cognition and Robotic Weapons Systems Under the International Law of Armed Conflict', 4 *Amsterdam Law Forum* (2012) 3–20.

one may well ask whether also the future deployment of fully autonomous weapons systems can be confined to situations where an encounter with the civilian population is excluded from the outset and where the actual decision with regard to the rules of international humanitarian law will be taken by the commander who sends the machine into the field. Indeed, given the prevailing forms of conflict, how realistic will such a scenario be? Conflicts are increasingly non-international, without clearly defined geographical front lines, with military targets located in predominantly civilian areas and with combatants that deliberately do not clearly distinguish themselves from non-combatants. As noted by one commentator, 'once autonomous weapons systems come into existence and become operational it will be difficult to avoid succumbing to the temptation to deploy them, even in complex and unpredictable situations'.[142] In fact, it is precisely in such complex situations that human soldiers are particularly under threat, making the incentive to replace them with an AWS particularly strong.

Accountability for breaches of international humanitarian law?

Autonomous weapons systems can be seen to pose a grave challenge to the notion of human accountability that lies at the heart of international humanitarian law as well as that of the system of international criminal justice that has developed to support and enforce it. The vexed question is: if fully autonomous systems will replace the role of humans in warfare, then where will accountability lie when something goes wrong?

Front-line operators?

When weapons systems are given increasing autonomy in order to respond to threats which humans may be too slow to respond to, in ways which are too complex for humans to control, it may become unrealistic to expect human operators to exercise significant veto control over their

142 Geiß, *The International–Law Dimension of AWS*, *supra* note 134 at 16.

operations. This is especially so given what is already known of the way in which humans interact with even relatively basic, semi-automated systems. That human operators may tend towards what is known as 'automation bias',[143] is well illustrated by the shooting down of Iran Air Flight 655 by the US Navy ship *USS Vincennes* in 1988.[144]

The *USS Vincennes* utilized a computer system which was designed to defend ships against airborne attacks. The system could be set to a number of different modes, each with varying degrees of automation, but in every mode the human operators had the ability to override the computer. On 3 July 1988, the radars of the system that was in semi-automatic mode detected Iran Air Flight 655. Albeit that the course, speed, radar broadcast, and radio signal coming from the plane all indicated that it was a commercial civilian flight, the ship's computer system registered the plane as an F-14 Fighter – an aircraft half the size of Flight 655. While all of the evidence should have made it clear that 655 was a civilian plane, no-one on board the *USS Vincennes* was willing to challenge the computer's judgment. The operators trusted the automated system although they knew that it had been specifically designed for engaging unfriendly Soviet bombers in the open North Atlantic, not for dealing with civilian-filled skies in the crowded Gulf. They authorized it to fire. Regardless of the wealth of evidence that outside of the automated system clearly indicated the actual nature of the target, the operators came thus to shoot down a civilian plane, killing all 290 passengers in one of the worst aviation incidents in history.[145] That the training of operators to trust the software leaves little room for human judgment and control, became also evident in the three incidents of fratricide that occurred during the so-called Operation Iraqi Freedom

143 See Mary L. Cummings, 'Automation and Accountability in Decision Support System Interface Design', 32 *Journal of Technology Studies* (2006) 26; and Peter Asaro, 'Modeling the Moral User', 28 *IEEE Technology and Society* (2009) 20.

144 For facts, see *Aerial Incident of 3 July 1988 (Islamic Republic of Iran* v. *United States) (Order of 22 February 1996)* ICJ Reports (1996) 9; see also Singer, *Wired for War*, *supra* note 2, at 124; and Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 14–15.

145 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 15. See also United States Department of Defence, *Investigation Report: Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988* (Department of Defence: Washington, DC, 1988) esp. Ch. IV, at 63.

in 2013: on two occasions, the electronic 'Patriot' missile defence system of the US forces engaged friendly coalition aircraft with deadly results, and in a third incident a US aircraft fired on a Patriot battery in the belief that it was an Iraqi missile system.[146]

Indeed, even though human operators are currently kept 'in the loop', their role is more often than not reduced to mere veto power – and even that is a power that operators are usually unwilling to use against the quicker (and what is most often viewed as better) judgment of a computer.[147] And who can blame them? It is only natural, particularly in complex situations, that an operator would be cautious about second guessing the computer's wisdom and running the risk of not preventing a potentially deadly attack. In situations like this, it would thus seem unfair to hold an individual operator accountable for failing to hit the 'off' switch. The deficit of human accountability has in such scenarios been qualified by many commentators more systemic than individual.[148]

Now, if automation bias prevails even with regard to relatively basic, semi-automated systems, then it is even more likely to do so in cases of advanced and more truly autonomous systems which make their own independent and complex decisions.

Computer programmers?

If front-line operators cannot be held responsible for the actions of autonomous weapons systems, it has been suggested that accountability could shift to the engineers, computer programmers, and designers who are

146 See Dan Saxon, 'A Human Touch: autonomous weapons, DoD Directive 3000.09 and the interpretation of "appropriate levels of human judgment over the use of force"' in Nehal Bhuta et al. eds, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 185–208 at 190–192.

147 Singer, *Wired for War*, *supra* note 2, at 124.

148 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 15.

responsible for creating the system in the first place.[149] As critics have been quick to point out, however, this would essentially result in a shift from the framework of international humanitarian law to a product liability framework, which arguably 'does not provide a strong enough sanction given the importance of and high stakes involved in enforcing compliance with international humanitarian law'.[150] Within the product liability framework, even those suspected of most heinous of war crimes would not have to face a criminal prosecution and a potential conviction but would be confined to a lawsuit and a potential monetary fine (that an insurance company would likely take care of). The product liability framework would also shift the burden of proof on the victims. How likely is the scenario, one may ask, that the civilian victims of an autonomous weapons system – poverty-stricken and geographically displaced by the conflict as they are certain to be – will sue for monetary relief against manufacturers in a foreign court?

Replacing the responsibility of a soldier with the responsibility of a computer technician is not straightforward. Criminal accountability is predicated on a guilty state of mind, and – even assuming that criminal liability for manufacturing errors could and should attach – the nature and complexity of AWS would make it very difficult indeed to successfully attach personal liability to any one programmer or engineer. As the weighing of collateral damage and military advantage requires the evaluation of multiple factors, a complete understanding of the risks associated with AWS may be impossible. While the balancing of factors would involve complex programming, it may never become possible to predict outcomes with any certainty.[151] Moreover, complex software is written not by one programmer, but by hundreds. Unforeseen interactions of code may result in undesirable re-

149 John F. Murphy, 'Mission Impossible? International law and the changing character of war', 41 *Israel Yearbook on Human Rights* (2011) 1–30 at 28; Brendan Gogarty & Meredith Hagger, 'The Laws of Man over Vehicles Unmanned: The Legal Response to Robotic Revolution on Sea, Land and Air', 19 *Journal of Law, Information and Science* (2008) 73 at 123.

150 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 16; see also, e.g., Gogarty & Hagger, 'The Laws of Man over Vehicles Unmanned', *supra* note 122 at 123; Stephen E. White, 'Brave New World: Neurowarfare and the Limits of International Humanitarian Law', 41 *Cornell International Law Journal* (2008) 177–210 at 209.

151 Marchant et al., 'International Governance', *supra* note 41 at 284.

sults, especially because AWS will be deployed in open and unstructured environments. For these reasons, it is unlikely that any court would impose liability on (let alone attribute responsibility to) a computer programmer whose small piece of code – possibly designed for much more general purposes than being used in an AWS – caused an unforeseen conflict within a massive code library, resulting in an AWS that acts in unpredictable or dangerous ways.[152]

Again, it has been suggested that liability may in this regard be more systemic than individual. For example, if a software company sells a piece of coding without properly testing it, and the military purchases and fields it knowing this, then who should be responsible? Militaries are suspected of quite regularly putting new weapons systems into production and use before they have been properly tested.[153]

The Reasonable Commander?

Facing the seemingly insurmountable challenges in the programming of an AWS, one commentator has concluded that these challenges may render their use 'almost useless except in the narrowest of circumstances'.[154] If AWS cannot be programmed to meet the 'reasonable commander' requirement in weighing the potential collateral damage and the military advantage, then they will never be capable of a proportional attack. To address this issue, another commentator has suggested that adhering to the principles of proportionality is really a question of probabilities: 'If the probability of success is low, or the probability of excessive collateral damage is high, then the weapon system will not engage'.[155] If AWS are operating in a civilian centre, the commander must set the threshold for engagement higher than if they were operating in a desert. Provided that the commander has programmed the AWS correctly, their use could be

152 Gogarty & Hagger, 'The Laws of Man over Vehicles Unmanned', *supra* note 122 at 123; White, 'Brave New World', *supra* note 123 at 209.

153 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 17.

154 Wagner, 'Taking Humans Out of the Loop', *supra* note 38, at 163.

155 Guetlin, *Lethal Autonomous Weapons*, *supra* note 49, at 11.

argued to be proportional.[156] This would put control of proportionality back in the hands of a human.

Indeed, yet another purported means of establishing accountability for the use and actions of autonomous weapons systems is the doctrine of command responsibility. The elements of command responsibility are set out in Article 28 of the Rome Statute of the International Criminal Court,[157] according to which:

> *In addition to other grounds of criminal responsibility under this Statute for crimes within the jurisdiction of the Court:*
>
> *(a) A military commander or person effectively acting as a military commander shall be criminally responsible for crimes within the jurisdiction of the Court committed by forces under his or her effective command and control, or effective authority and control as the case may be, as a result of his or her failure to exercise control properly over such forces, where:*
>
> *(i) That military commander or person either knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes; and*
>
> *(ii) That military commander or person failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter to the competent authorities for investigation and prosecution.*
>
> *(b) With respect to superior and subordinate relationships*

156 Guetlin, *Lethal Autonomous Weapons*, *supra* note 49, at 12.

157 Article 28 of the Rome Statute is considered to codify existing customary international law, see Knut Dörmann, 'War Crimes under the Rome Statute of the International Criminal Court, with a Special Focus on the Negotiations on the Elements of Crimes', 7 *Max Plank Yearbook of United Nations Law* (2003) 341 at 345.

not described in paragraph (a), a superior shall be criminally responsible for crimes within the jurisdiction of the Court committed by subordinates under his or her effective authority and control, as a result of his or her failure to exercise control properly over such subordinates, where:

(i) The superior either knew, or consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit such crimes;

(ii) The crimes concerned activities that were within the effective responsibility and control of the superior; and

(iii) The superior failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter to the competent authorities for investigation and prosecution.

Command responsibility as it is presently understood clearly concerns the responsibility of a commander for the *people* under his or her command.[158] It has been suggested, however, that the concept could be expanded to cover responsibility for autonomous weapons systems too.[159] The proposal can be regarded sensible in that it addresses head-on the unique aspect of AWS which makes them so worrisome: that *rather than just being another kind of weapon, they are really replacing the human role of the soldier*.[160]

Indeed, against the background of the ever-prevalent 'automation bias', it does seem to make sense to place responsibility for AWS at a relatively

158 Murphy, 'Mission Impossible?', *supra* note 122 at 27.

159 White, 'Brave New World', *supra* note 123 at 205.

160 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 18; see also Neha Jain, 'Autonomous weapons systems: new frameworks for individual responsibility' in in Nehal Bhuta et al. eds, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 303–324 at 303.

high level. As the Director for International Law and Cooperation at the ICRC has commented in this regard, someone has to be responsible for turning the system on.[161] It may be that some difficulties in accountability could be avoided if that 'someone' was a relatively high-level commander.

However, it does also seem unlikely that in the context of AWS the accuracy and reliability of the weapons could be measured in the usual ways. Whereas, for example, the accuracy of rockets and bombs can be measured in terms of a radius – the legitimacy of using the weapon then becoming dependent, in part, on the number of civilian objects within that radius – the accuracy of an AWS might, in contrast, depend on such complex factors as how many objects the system will have to analyse, the ability of its sensors to properly analyse the kinds of objects in range, and whether it is operating in an area where there is greater scope for mistaken targets. This makes the determination of reliability or accuracy difficult. Moreover, any such information may be confidential and not available at every level of the military.[162]

It has been argued that this can, in a sense, be seen a round-about way of attaching criminal liability to a failure to comply with the requirement to take precautions in attack.[163] In particular, if a commander *should* have known that an autonomous weapon was likely to commit the equivalent of a war crime but turned it on anyway, then (s)he has not taken 'all feasible precautions in the choice of means and methods of attack with a view to avoiding . . . incidental loss of civilian life'[164] or refrained 'from deciding to launch any attack which may be expected to cause incidental loss of civilian life . . . which would be excessive in relation to the concrete and direct military advantage anticipated'.[165]

However, in order to apply either the doctrine of command responsibility or the duty to take precautions in the context of AWS, the system should

161 Dr Philip Spoerri, Director for International Law and Cooperation, ICRC, 'Round Table on New Weapon Technologies and IHL – Conclusions', *34th Round Table on Current Issues of International Humanitarian Law*, 8–10 September 2011, San Remo.

162 White, 'Brave New World', *supra* note 123 at 206–207.

163 Grut, 'Challenge of Autonomous Lethal Robotics', *supra* note 68, at 19.

164 Cf. Protocol I, *supra* note 27, Art 57; *ICRC Study*, *supra* note 67, Rule 15.

165 Cf. Protocol I, *supra* note 27, Art 57; *ICRC Study*, *supra* note 67, Rule 15.

have at least some degree of predictability.[166] Otherwise the commander can hardly be regarded to have been in a position where she or he 'should have known' that the system was committing or about to commit a 'crime': it could not be 'expected' that the use of this particular means of warfare would cause excessive incidental loss. Particularly with more complex systems, it may not always be the case that the AWS is predictable enough to make even this kind of human accountability realistic or sensible: 'programs with millions of lines of code are written by teams of programmers, none of whom knows the entire program; hence, no individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways'.[167]

As autonomous weapons systems with strong artificial intelligence are expected to be capable of independent learning, there is also a genuine concern as to whether it will even be possible to program AWS so that they will obey orders.[168]

166 Albeit that the command responsibility under Article 28 of the Rome Statute considerably lowers the *mens rea* requirement when compared to that required for perpetration or even complicity in international law; this in turn broadens the potential scope of liability to cover individual cases of AWS deployment gone awry; see Jain, 'Autonomous weapons systems', *supra* note 160 at 315 and 319. Cf., however, Darryl Robinson, 'How Command Responsibility Got So Complicated: A Culpability Contradiction, Its Obfuscation, and a Simple Solution', 13 *Melbourne Journal of International Law* (2012) 1–58; and Ilias Bantekas, 'The Contemporary Law of Superior Responsibility', 93 *American Journal of International Law* (1999) 573–595.

167 Marchant et al., 'International Governance', *supra* note 41 at 283.

168 See Sparrow, 'Killer Robots', *supra* note 6; Marchant et al., 'International Governance', *supra* note 41 at 285.

4 THE EFFECT OF AUTONOMOUS SYSTEMS ON HUMAN RIGHTS IN ARMED CONFLICT

While the legal debate about autonomous weapons systems has during the past few years primarily focused on international humanitarian law, some of the more critical voices have also raised specific concerns as to the permissibility of autonomous weapons under international human rights law.[169] To be sure, human rights law is understood to complement the rules of international humanitarian law during armed conflict.[170] This complementarity is subject to a number of qualifications, however, including the proviso that during armed conflict the rules of human rights law are determined with reference to the provisions of international humanitarian law which are more specialized – and in many ways more permissive.[171] Yet, importantly, people on both (all) sides of the conflict retain their human rights such as the right to life and the right to dignity during armed conflict, albeit that the more specific contents of the rights may differ according to the context.[172] The rules of international humanitarian law should, in turn, be interpreted with reference to these rights. However, as noted by the United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions, Professor Heyns, the 'cumulative effect' of international human rights standards leads to a 'fundamental incompatibility' between the constitutive values of the human rights regime and the use of

169 See, especially, HRW & IHRC, *Losing Humanity*, *supra* note 63; HRW & IHRC, *Shaking the Foundations: The Human Rights Implications of Killer Robots* (12 May 2014) available at <www.hrw.org/report/2014/05/12/shaking-foundations/human-rights-implications-killer-robots>; and HRW & IHRC, *Killer Robots and the Concept of Meaningful Human Control: Memorandum to Convention on Conventional Weapons (CCW) Delegates, April 2016* (11 April 2016) available at <www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control>.

170 See, e.g., Silvia Borelli & Simon Olleson, 'Obligations Relating to Human Rights and Humanitarian Law' in James Crawford, Alain Pellet & Simon Olleson (eds), *The Law of International Responsibility* (Oxford University Press 2010) 1177–1198.

171 See Françoise J. Hampson, 'The relationship between international humanitarian law and human rights law from the perspective of a human rights treaty body', 90 *International Review of the Red Cross* No. 871 (2008) 549–572.

172 See Cordula Droege, 'Elective Affinities? Human Rights and Humanitarian Law', 90 *International Review of the Red Cross* No. 871 (2008) 501.

autonomous weapons systems.[173] Indeed, there is considerably less space for the use of autonomous weapons under human rights law than there is under international humanitarian law.

Those human rights that critics highlight as being most notably infringed by the potential use of autonomous weapons include the right to life and the right to human dignity. As for right to life,[174] international human rights law poses a number of rules for the use of force which have titles similar to those used in international humanitarian law, but which differ greatly in their content. This includes the rules of 'necessity', which in the human rights context means that force should only be used as a last resort, and if that is the case, a graduated approach should be followed. While the hostile intention of the target is irrelevant in the context of international humanitarian law, where the focus is on status or conduct, it often plays a decisive role in the human rights context. 'Proportionality', for its part, sets in this context a maximum on the force that may be used to achieve a specific legitimate purpose: the interest harmed may not exceed the interest protected. The fact that force may be 'necessary' does not imply that it is proportionate. Moreover, the argument that a deadly return of fire is justified as self-defence, is not deemed available within human rights framework insofar as autonomous weapons systems are concerned: 'Intentional deadly force may be used only to protect human life, and not objects such as a machine'.[175]

The requirements for the use of force under human rights law are clearly much stricter than under international humanitarian law. A case-by-case assessment is needed, not only of each attack, as under international humanitarian law, but of each use of force against a particular individual. The problem encountered in the context of armed conflict is whether ma-

173 See Christof Heyns, 'Autonomous Weapons Systems and Human Rights Law', Presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons 13 – 16 May 2014, Geneva, Switzerland, available at <www.icla.up.ac.za/images/un/speeches/heyns ccw presentation aws and human rights.pdf> at 5.

174 See United Nations Human Rights Committee, *CCPR General Comment No. 6: Article 6 (Right to Life), Adopted at the Sixteenth Session of the Human Rights Committee, on 30 April 1982*, available at <www.refworld.org/docid/45388400a.html>.

175 Heyns, 'Autonomous Weapons Systems and Human Rights Law', *supra* note 149 at 6.

chines have, or will ever have, the ability to make the qualitative assessments required for the use of force. In assessing the potential effect of autonomous weapons on human rights, UN Special Rapporteur Heyns has emphasized that it is very difficult to conceive that machines will be able to ascertain whether a particular person has the intention to attack with sufficient certainty to warrant the release of deadly force: 'Allowing machines to determine whether to act in defence of others poses grave risks that the right to life will be violated'.[176] He has also argued that a determination of life and death by a machine is 'inherently arbitrary'. In his view, there is an unspoken assumption in international human rights law that the final decision to use lethal force must be reasonable and taken by a human.[177] As machines cannot 'reason' in the way that humans do, they cannot therefore take 'reasonable' decisions on their own.

Many have argued that the notion of the right to life cannot be understood in isolation from the concept of dignity, because it is the value of life that makes it worth protecting.[178] Such 'right to dignity' is perceived to be at the heart of the international human rights canon and, arguably, also in the rules of international humanitarian law.[179] Certainly, the so-called Martens Clause, in its traditional as well as in its modern occurrence in the Additional Protocol I to the Geneva Conventions relating to the protection of victims of international conflicts, by referring to the principles of humanity and the dictates of public conscience, allows for the consideration

176 Heyns, 'Autonomous Weapons Systems and Human Rights Law', *supra* note 145 at 6.

177 See, e.g., Heyns, *Report of the Special Rapporteur*, *infra* note 65 at para. 89.

178 See Bernhard Schlink, 'The Concept of Human Dignity: current usages, future discourses' in Christopher McCrudden ed., *Understanding Human Dignity* (Oxford University Press, 2013) 632; Paolo G. Carozza, 'Human Dignity and Judicial Interpretation of Human Rights: a reply', 19 *European Journal of International Law* (2008) 931–44; and Anne Phillips, *The Politics of the Human* (Cambridge University Press, 2015) chapter 4.

179 Article 1 of the Universal Declaration of Human Rights provides as follows: 'All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood'. While dignity is not recognized as a separate right in the International Covenant on Civil and Political Rights, it is considered a constitutive part of a number of the rights contained in that treaty. It is also recognized in several treaties as a separate right and it is a concept that influences the way in which other rights are interpreted.

of human dignity for the purposes of the law of armed conflict.[180] In the context of the potential use of autonomous weapons systems, the right to dignity serves primarily to protect those targeted, rather than those who are incidental casualties. It is because of this, that critics from the human rights field have sought to draw attention to the fact that the almost exclusive emphasis in much of the current legal debate about the effect of autonomous weapons on civilian casualties and their right to life has skewed the discussions: a significant but under-emphasized part of the problem with autonomous weapons is their potential impact on the very dignity of those targeted.[181] In the similar vein, it has been argued that having a machine decide whether one lives or dies is a question of 'ultimate indignity'.[182] This argument has been extended by some to the decision by machines to use force in general.[183] At the root of the argument is the idea that 'death by algorithm' objectifies people, treating an individual as an interchangeable entity, rather than a human being with inherent dignity. A decision as far-reaching as to deploy deadly force should, it is emphasized, only be taken by a human being, after due consideration: 'A machine, bloodless and without morality or mortality, cannot fathom the significance of the killing or maiming of a human being'.[184] It has also been put forth that as dignity in many instances depends on hope, all dignity is stripped off by the very knowledge of the fact that one may at any moment be confronted by an autonomous weapon system which will bring one's life to an end 'with all the certainty that science can offer' thus leaving 'no room for the

180 According to Article 1(2) of Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, *infra* note 27: '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 dictates of public conscience.'

181 See Dieter Birnbacher, 'Are autonomous weapons systems a threat to human dignity?' in Nehal Bhuta et al. eds, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 105–121.

182 See Christof Heyns, 'Autonomous weapons systems: living a dignified life and dying a dignified death' in Nehal Bhuta et al. eds, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 3–20.

183 See Birnbacher, 'Are autonomous weapons systems a threat', *supra* note 153.

184 Heyns, 'Autonomous Weapons Systems and Human Rights Law', *supra* note 145 at 8. See also Birnbacher, 'Are autonomous weapons systems a threat', *supra* note 153.

possibility of an exception; for a rare occurrence of compassion or just a last-minute change of mind'.[185]

* * *

The concerns that states now have over weapons with the power to make life-and-death determinations have since 2014 been discussed by the states parties to the Convention on Conventional Weapons (CCW)[186] at meetings devoted to 'lethal autonomous weapons systems', which is the chosen CCW term for fully autonomous weapons.[187] During the discussions, it has become apparent that many of the states are adamantly opposed to assessing the legality of the means of warfare in terms of human rights. That this is so should have come as no surprise, as it is well known based on the experiences in the United Nations Human Rights Council and before the European Court of Human Rights, that states such as the United Kingdom, United States and Israel, for example, stand against any consideration of the means and methods of warfare in light of substantive human rights.[188] Thus, both the United States and the United Kingdom have also during the recent CCW rounds of experts in Geneva pointed out explicitly that international humanitarian law is the only correct framework for assessing autonomous weapons systems and that this narrow focus corresponds to the mandate of the CCW regime.[189]

185 Heyns, 'Autonomous Weapons Systems and Human Rights Law', *supra* note 145 at 8.

186 Cf. UN Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, *supra* note 129.

187 For background, see the United Nations Office at Geneva website on autonomous weapons systems within the CCW framework, <www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument>

188 Bhuta, Beck & Geiß, 'Present Futures', *infra* note 218 at 381.

189 See U.S. Delegation Opening Statement As Delivered by Michael W. Meier, The Convention on Certain Conventional Weapons (CCW) Informal Meeting of Experts on Lethal Autonomous Weapons Systems, 13 April 2015, Geneva, available at <www.unog.ch/80256EDD006B8954/(httpAssets)/8B33A1CDBE80EC60C1257E2800275E56/$file/2015_LAWS_MX_USA+bis.pdf> and United Kingdom of Great Britain and Northern Ireland Statement to the Informal Meeting of Experts on Autonomous Weapons Systems, 13 – 17 April 2015, <www.unog.ch/80256EDD006B8954/(httpAssets)/1CBF996AF7AD10E2C1257E260060318A/$file/2015_LAWS_MX_United+Kingdom.pdf>.

During the discussions, it has become equally evident that a more comprehensive ban on autonomous technology in military systems now seems unlikely and indeed unrealistic. Thus far, only Bolivia, Cuba, Ecuador, Egypt, Ghana, Pakistan, and Zimbabwe as well as the Holy See and Palestine have publicly endorsed a comprehensive ban[190] – a treaty concluded between only a few like-minded states but without the participation of states with high-tech military capacities, would be of little avail. The anticipated advantages and strategic promises of autonomous weapons systems are such that high-tech states can be expected to agree to certain limitations on further development of autonomous capabilities only if all the other competing high-tech states do the same. At the current juncture, two main areas have emerged under the aegis of the CCW as central points for discussion, namely enhanced weapons reviews and the notion of 'meaningful human control'. The present research report will also engage with these two areas as potential starting points for a future international regulation of autonomous weapons technology.

190 Cf. Campaign to Stop Killer Robots, *Report on Activities: Convention on Conventional Weapons Annual Meeting of High Contracting Parties, United Nations Geneva, 12–13 November 2015* (Washington, DC, 16 December 2015) at 4–5, available at <www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_ReportCCWannual16Dec2015_uploaded-1.pdf >.

5 POSSIBLE WAYS TO REGULATE AUTONOMOUS WEAPONS SYSTEMS

Historical experience suggests that if a new weapon has substantial advantages for one state, then it will gradually be adopted by other states too.[191] Given the considerable advantages that autonomous weapons systems potentially offer, it is very likely indeed that their use will proliferate already in the next decade or so. In what follows, I shall briefly present the most important contemporary proposals for the containment of AWS.

Requirement of 'Meaningful Human Control'

The requirement of 'meaningful human control' over weapons has recently emerged as a substantively open and therefore constructive starting point for the debates on the acceptability of autonomous weapons. It appears from the expert meetings on AWS, held by the states parties to the Convention on Conventional Weapons since 2014, that regardless of whether they are against or in favour of increasing autonomy, there is a consensus among various states and experts that a certain degree of human control is instrumental to the public acceptance of AWS. More specifically the consensus seems to be that AWS should be permitted only to the extent that meaningful human control is retained in relation to the most critical of decisions, above all the decision to deploy lethal force.[192] This demand has also been supported by international and non-governmental organizations such as the International Committee of the Red Cross,[193] the

191 Kenneth Anderson & Matthew C. Waxman, 'Law and Ethics for Robot Soldiers', 176 *Policy Review* (2012) 35–49 at 40; see also Anderson & Waxman, 'Law and Ethics for Autonomous Weapons Systems: Why a Ban Won't Work and How the Laws of War Can' (Hoover Institution, Stanford University 2013) at 8, available at <www.hoover.org/research-teams/national-security-law-task-force>.

192 See, e.g., remarks by Thomas Nash, Director, Article 36, Informal Expert Meeting on Lethal Autonomous Weapon Systems, Convention on Certain Convention of Weapons, Geneva (2014).

193 Statement by the International Committee of the Red Cross (13 May 2014) available at <www.unog.ch/80256EDD006B8954/(httpAssets)/C99C06D328117A11C1257CD7005D8753/$file/ICRC_MX_LAWS_2014.pdf>.

Campaign to Stop Killer Robots,[194] and the International Committee for Robot Arms Control.[195]

Indeed, and as is evident also from the preceding chapters of the present research report, it is very likely that without any human control some targeting decision by AWS in armed conflicts cannot be made in compliance with international humanitarian law (such as proportionality decisions, say). Moreover, a responsibility vacuum is likely to ensue. But exactly how the requirement of 'meaningful' human control might alleviate such situations, remains a wholly open question, subject to debate.

Ultimately, of course, the criterion does translate to a ban on full autonomy over certain critical functions of a weapons system – for if there is human control, by definition there cannot be full autonomy.[196] It may thus seem somewhat paradoxical to evoke the criterion of meaningful human control in the context of 'fully autonomous' weapons systems. And yet, quite like the paradoxical notion of 'sustainable development' that for decades now has served as a useful platform for bringing together and striking compromises between conflicting views within international environmental law, the requirement of meaningful human control over autonomous weapons systems now seems to offer a useful and constructive starting point for further substantive discussions on how to contain the foreseeable proliferation of AWS.

From here on, then, the international debate is likely to focus on defining the critical decisions in the deployment of AWS that should always be subject to meaningful human control and on precisely how such control should be implemented across the spectrum of autonomy.[197] In this, the required level of control may refer to several factors: to the time-span between the last decision taken by a human and the exertion of force by the AWS; to the environment in which the AWS comes to be deployed,

194 Campaign to Stop Killer Robots statement by Mary Wareham, Human Rights Watch (16 May 2014) available at <www.unog.ch/80256EDD006B8954/(httpAssets)/16B608BD428C6D17C1257CDA0056AA62/$file/NGO+Campaign+StopKillBots_FinalStatement.pdf>.

195 Statement by the International Committee for Robot Arms Control (13 May 2014) available at <www.unog.ch/80256EDD006B8954/(httpAssets)/8A68157979FEFBB6C1257CD7006AB5FD/$file/NGO+ICRAC+MX+LAWS.pdf>.

196 Bhuta, Beck & Geiß, 'Present Futures', *infra* note 224 at 381.

197 Cf. Scharre & Horowitz, *An Introduction to Autonomy*, *supra* note 5.

especially as to the question of whether civilians are potentially present in that environment; to the nature of the deployment, that is to say, whether the AWS is supposed to engage in a defensive or an offensive mission; to the question of whether the AWS is designed or set up to apply lethal force; to the quality and the level of training of the operators that are to exercise control over the AWS; to the question of the extent to which a human is in a position to intervene, should the need arise, and halt the mission; and to the implementation of safeguards with regard to responsibility by, say, making sure there is an electronic recording of all the actions of the AWS.[198]

The field of artificial intelligence is advancing rapidly, and it is conceivable that in the future autonomous weapons will be able to learn independently and modify their pre-programmed rules of conduct in response to experiences or changes in their environment. If humans can no longer predict how an autonomous weapon will behave, then meaningful human control will cease to exist. After all, such control can only exist if humans are able to anticipate how an autonomous weapon will behave in a particular situation and if they are able to retroactively explain its behaviour. If it becomes impossible to predict how an autonomous weapon will behave, then it is debatable whether a commander can still make a judicious decision concerning its deployment, as he or she cannot be certain that laws of war will not be violated. Another scenario for meaningful human control being partially or largely wiped out by technological advances involves the full or partial transfer of the command-and-control function to computers that could accidentally activate autonomous weapons. Also, the increasing complexity of autonomous systems could ultimately lead to a partial or near-complete loss of human control. If this happens, it is impossible to predict whether technology will find ways to guarantee that autonomous weapons continue to function in accordance with international legal norms and ethical principles.

Currently, the benefits of removing humans from the decision-making loop are minimal: the technology is not sophisticated enough to replace a human. As the technology improves, however, the benefits of removing

198 Geiß, *The International-Law Dimension of AWS*, *supra* note 138 at 24–25; AIV & CAVV, *Autonomous Weapons Systems*, *supra* note 24 at 34–39.

humans will grow. To be sure, experts have already noted that while technological developments in the field of autonomous systems are incremental,[199] it is likely that human involvement in the selection and engagement of targets will gradually erode.[200] Ensuring that humans stay in the loop may perhaps be regarded a temporary solution to the adherence of AWS to the principles of international humanitarian law, but it is not a permanent solution – unless there is a legal framework that guarantees the role of meaningful human control at various stages of the targeting process.

Constraining Through Programming: an 'Ethical Governor'

A number of computer scientists have proposed that it will be possible to embed ethics in AWS in order to ensure that their behaviour adheres to the principles of international humanitarian law. By constraining the range of possible actions available to AWS through an algorithm that is to act as an 'ethical governor' and 'ethical behavioural control', they contend that AWS will not only be capable of adhering to international humanitarian law, but that they will also be able to exceed human capabilities in doing so.[201] Such an 'ethical' algorithm would embed a feedback loop into the control software of the AWS that would then either allow the system to deploy its weapons in a particular instance or forbid it from doing so.[202] A review authority of sorts would thus be programmed into the AWS that would evaluate the compliance of the proposed action with international humanitarian law prior to any actual employment of lethal force. Although experts do not rule out the possibility that even with such an 'ethical

199 See Krishnan, *Killer Robots*, *supra* note 35; see also Alex Leveringhaus and Gianni Giacca, *Robo-Wars: The Regulation of Robotic Weapons* (Oxford Martin Policy Paper, 2014).

200 See 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* (2013) at 19, available at <papers.ssrn.com/sol3/papers. cfm?abstract_id=2250126>.

201 Ronald C. Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture* (Technical Report GIT-GVU-07-11, Georgia Institute of Technology, 2007) at 61; see also Krishnan, *Killer Robots*, *supra* note 35.

202 Arkin, *Governing Lethal Behavior*, *supra* note 202.

governor' AWS may make mistakes, they do assume that the margin of error for an AWS would be significantly smaller than for a human soldier.

The attempts to translate the principles of international humanitarian law into a logical and programmable structure may indeed offer a practical solution to some of the issues raised by AWS, but there are problems with this approach. Such an 'ethical governor' assumes that there are 'effective situational assessment methods' in place to ensure that an AWS would not commit a lethal mistake.[203] As demonstrated above, this assumption may not be appropriate.

The Obligation to Review and Codes of Conduct

The inspection mechanisms under international humanitarian law for the introduction of new weapons are, in themselves, applicable also to autonomous weapons systems. Article 36 of Additional Protocol I requires that each member state 'is under an obligation to determine whether [a new weapon's] employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law'.[204] It has been suggested that even states, such as the United States, that are not a signatory to this Protocol are obligated to review weapons and ensure that they are capable of complying with the principles of international humanitarian law.[205]

To be sure, for example the US Department of Defence has confirmed that 'the acquisition and procurement of weapons and weapon systems shall be consistent with all applicable domestic law and treaties and international agreements, . . . customary international law, and the law of armed conflict'.[206] States engaged in the development of AWS can be seen

203 Krishnan, *Killer Robots*, *supra* note 35, at 22, 46, 109.

204 Article 36 of Additional Protocol I to the Geneva Conventions, *supra* note 27.

205 See Schmitt, 'Autonomous Weapons Systems', *supra* note 7 at 28.

206 United States Department of Defence, Directive Number 5000.01: The Defense Acquisition System, 20 November 2007, available at <www.dtic.mil/whs/directives/corres/pdf/500001p.pdf>.

to have an inherent interest in the regulation of the AWS.[207] The concern that, for example, the United States has about the development of AWS is demonstrated in the 2012 Department of Defence Directive that represents the current (public) US code of conduct.[208] Apart from the general concern, however, the Directive does not in any way *plan* for the future development of AWS. Instead, as noted, it seemingly relies on keeping humans in the loop. While this may be an effective way of ensuring the legality of AWS in the short term, it does not prepare for the possibility that when it becomes more advantageous and technologically plausible to take humans out the loop, a new policy will have to be implemented to ensure compliance with international humanitarian law. Such an approach may result in internal safeguards only, without promoting international discussion nor ensuring wider international compliance with the principles of international humanitarian law.

There are many benefits to a 'code of conduct' approach. Codes are more flexible than formal multilateral conventions, and are potentially more capable of adapting quickly to technological advances. They are also relatively easy to create and adopt, especially so in comparison to multilateral conventions. However, there are disadvantages too. National codes of conduct are not internationally binding, and do not involve any international oversight. Thus, although they may offer a reliable short-term solution and aid in shaping international dialogue, they do not offer solutions in the long term, as more and more states become capable of developing and using AWS. While some states may adopt stricter policies as humans are moved further from the AWS loop, others may not. Relying on state-created codes of conduct and their obligation to review will result in unilateral decision-making. This would frustrate the development of internationally recognized principles necessary to ensure the safe operation of AWS on a global scale.

Against this background, various states have now availed themselves of the occasions of the Geneva expert meetings on AWS to advocate further development and specification of the substantive requirements under Ar-

207 See, e.g., Anderson and Waxman proposing that the United States develop a set of principles to regulate and govern AWS, not only for themselves, but for the benefit of the international community as a whole, in 'Robot Soldiers', *supra* note 161 at 46.

208 See US DoD, 'Autonomy in Weapon Systems', *supra* note 34.

ticle 36, especially with regard to autonomous weapons systems.[209] More detailed elaboration of the Article 36 procedures would undoubtedly be helpful to ensure a more widespread implementation thereof. As Article 36 calls for determinations of a new weapon's compliance with international humanitarian law in the study, development and acquisition phases of its production, this has been interpreted to mean that for a state producing weapons, 'reviews should take place at the stage of conception/design of the weapon, and thereafter at the stages of its technological development (development of prototypes and testing), and . . . before entering into the production contract'.[210] At least two legal reviews, then, are required: one before taking the decision to begin the formal development of the AWS, and another before the AWS is fielded. Tests carried out by manufacturers and potential buyers should be carried out in realistic environment so as to reveal how the weapon behaves under various conditions and to indicate potential risks associated with its deployment. Such tests can produce valuable information for commanders who will be responsible for deciding whether or not to deploy AWS.[211] Any such procedure for assessing the compatibility of AWS with Article 36 should also examine whether the degree to which human control has been incorporated into the design of the weapon offers adequate guarantees of compliance with international law.

Banning Autonomous Weapons Systems

A number of non-governmental organizations have been calling for autonomous weapons systems to be banned under an international agreement akin to that prohibiting anti-personnel landmines.[212] Human Rights Watch has been particularly active in this regard,[213] as has been the 'Campaign to

209 See Geiß, *The International-Law Dimension of AWS*, *supra* note 138 at 26.

210 International Committee of the Red Cross, *A Guide to the Legal Review of New weapons, Means and Methods of Warfare* (ICRC: Geneva, 2006) at 23.

211 AIV & CAVV, *Autonomous Weapons Systems*, *supra* note 24 at 36.

212 Cf. Convention on the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, *supra* note 32.

213 See, e.g., HRW & IHRC, *Losing Humanity*, *supra* note 70; HRW & IHRC, *Shaking the Foundations*, *supra* note 168; HRW & IHRC, *Killer Robots, supra* note 141.

Stop Robot Killers' initiative.[214] The proposal is to simply prohibit any use of lethal force by AWS without human input or supervision so as to effectively ensure that AWS adhered to the principles of international humanitarian law and so as to put an end to any further dehumanisation of war that would but lower the threshold for states to commence armed conflicts. Moreover, as the problem of the responsibility gap seems irresolvable, the argument is that the looming liability loopholes can only be countered by means of a total ban. According to Human Rights Watch these problems arise not just with regard to completely autonomous systems, but also in relation to weapons that have so much influence over the selection of targets that human controllers no longer exercise any real control and are thus effectively 'out of the loop'.[215] Some experts too have lent their support to a ban, arguing that there are ethical and legal duties to prevent autonomous weapons systems from ever being given the capability of deciding on human life and death.[216] Only a handful of states, however, have publicly endorsed such a comprehensive prohibition – namely Bolivia, Cuba, Ecuador, Egypt, Ghana and Pakistan, as well as Palestine and the Holy See.[217]

Opponents of the ban maintain that as there is a very real possibility – likelihood, even – that AWS will, as such, be capable of lawful use, it would be irresponsible to prohibit them especially at this stage of development.[218] In fact, assuming that there is also a very real possibility that AWS would be more capable of distinguishing and acting proportionally than human soldiers, their ban would only serve to increase the risks to civilians.[219] The argument thus is that if the potential advantages of AWS are accepted, an outright ban will be more, rather than less, harmful to the application of international humanitarian law. Moreover, and importantly, the opponents of the ban have pointed out the mere futility of a comprehensive ban of the technologies that the development of AWS relies on, for these are also

214 See the campaign website at <www.stopkillerrobots.org>.

215 HRW & IHRC, *Losing Humanity*, *supra* note 70 at 46.

216 See, e.g., Asaro, 'On banning AWS', *supra* note 109, at 689.

217 See *supra* note 188 and the accompanying text.

218 See, e.g., Rebecca Crootof, 'The Killer Robots Are Here: Legal and Policy Implications', 36 *Cardozo Law Review* (2015) 1837–1915.

219 Schmitt, 'Autonomous Weapons Systems', *supra* note 7, at 25.

the technologies of the hardware and the software that the everyday routines of our increasingly automated societies have already come to rely on.

Realistically, an outright ban would indeed be a step too far. The proposal for the ban has nevertheless enhanced the legal discourse on AWS, promoting a serious consideration of the issues faced in the implementation of autonomous systems.

Multilateral Conventions

Should an AWS under development seem essentially incapable of meeting the legal standards set by international humanitarian law, then it would be unlawful already from the outset, as much is clear. However, even if such a newly developed AWS could, in principle, be considered *capable of lawful use*, this alone might not be sufficient to ensure that the AWS would *actually be used lawfully* – neither could the applicable standards of international humanitarian law nor the potential criminal accountability ensuing from a breach of those standards guarantee with any certainty that the actual use of AWS would be lawful. This is a concern that is by no means limited to AWS.

For example, the widespread use of anti-personnel landmines and cluster munitions gave many states cause for concern that the principles of international humanitarian law were insufficient to protect civilians from their use. Although such weapons were, in principle, capable of lawful use, in reality they were seldom used in conformity with international humanitarian law. Accordingly, rather than relying solely on the principles of international humanitarian law, states responded to the challenges raised by anti-personnel landmines and cluster munitions through multilateral conventions.[220] There are in fact numerous examples of new weapons technology giving rise to multilateral conventions that restrain their use or

220 See Convention on the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, *supra* note 32.

development: biological weapons,[221] chemical weapons,[222] blinding lasers,[223] and landmines.[224] While these conventions have employed a wide range of approaches to regulating the use of weapons,[225] all of them have been considered successful in the regulation.[226] This, it has been said, is so because the use of the regulated weapons was so widely recognized as contrary to the principles of international humanitarian law.[227]

The same cannot be yet be said of autonomous weapons systems, for there is still reasonable debate as to whether or not AWS will be capable of adhering to international humanitarian law. Accordingly, it has been suggested that a multilateral treaty that bans or limits the use of AWS is impractical. Consensus would be impossible to reach, and it would be impossible to ensure compliance.[228] The discussions in the Geneva expert meetings on AWS demonstrate that further development of AWS may be necessary before a multilateral convention will be considered a viable proposal. And yet, waiting until the technology actually exists before starting to develop standards to govern its use, is not preferable either. The development of AWS that are capable of meeting international humanitarian law standards requires guidelines – and constant dialogue.

221 Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and or their Destruction, 10 April 1972, entered into force 26 March 1975, 1015 *United Nations Treaty Series* 163.

222 Convention on the Prohibition of the Development, Production and Stockpiling of Chemical Weapons and on their Destruction, 13 January 1993, entered into force 29 April 1997, 1974 *United Nations Treaty Series* 45.

223 Protocol on Blinding Laser Weapons, *supra* note 33.

224 Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices *supra* note 32.

225 For example, both the Biological Weapons Convention and the Chemical Weapons Convention prohibit all parties from acquiring, developing, producing, stockpiling, or retaining biological or chemical weapons; the Protocol IV on Blinding Laser Weapons on the other hand prohibits neither the development nor the research of permanently blinding laser weapons, instead it prohibits their use only.

226 See Marchant et al., 'International Governance', *supra* note 42 at 301.

227 James Foy, 'Autonomous Weapons Systems: Taking the Human Out of International Humanitarian Law', 23 *Dalhousie Journal of Legal Studies* (2014) 47–70 at 66.

228 See Anderson & Waxman, 'Robot Soldiers', *supra* note 161 at 45.

Framework Conventions

Yet another possibility for the regulation of autonomous weapons systems would be the adoption of a framework convention.[229] This approach would combine the regulative qualities of a multilateral convention with a more flexible 'code of conduct' approach. A bare-bones multilateral convention could be created that develops a process and institutional capacity to gradually develop a substantive legal regime. Examples of such a framework convention include the Vienna Convention for the Protection of the Ozone Layer[230] and the Framework Convention on Climate Change.[231] Indeed, as the experience from the wide and varied field of international environmental law suggests, a framework convention has several benefits.[232] By first acknowledging the existence of a problem, it draws the attention of experts and the public to it. And then, it will gradually begin to commit states to taking more substantive action in the future. In the specific instance of autonomous weapons systems, such a regime would not, at first, be a clearly defined regime from the start, but it would allow for the identification of the precise issues that need addressing. While it would not commit states to a binding agreement before the full (or, at least, fuller) capabilities of AWS are known, it would from the start provide a structured venue for an open dialogue among states that are in various stages of development of AWS. It would in fact allow for the inclusion of all stakeholders. Through an incremental approach, it would also enable the realistic implementation of international standards and periodic reassessment by providing guidelines and listing best practices on such issues as the role of meaningful human control as a benchmark for assessing compatibility with Article 36, whether in the design or the actual deployment of autonomous weapon systems.

229 See Marchant et al., 'International Governance', *supra* note 41 at 313.

230 Vienna Convention on the Protection of the Ozone Layer, 22 March 1985, entered into force 22 September 1988, 1513 *United Nations Treaty Series* 323.

231 United Nations Framework Convention on Climate Change, New York, 9 May 1992, entered into force 21 March 1994, 1771 *United Nations Treaty Series* 107.

232 See Jarna Petman, 'Deformalization of International Organizations Law' in Jan Klabbers & Åsa Wallendahl (eds), *Research Handbook on the Law of International Organizations* (Edward Elgar Publishing Ltd: Cheltenham 2011) 398–430 at 405–409.

To be sure, the international community cannot wait until there is a fully autonomous system in place before moving to resolve the legal questions involved. A legal framework must exist before the development of autonomous weapons has advanced so far that their underlying architecture is difficult, if not impossible, to change. Law cannot be an afterthought.

6 CONCLUSION

Autonomy has been an increasingly important feature of offensive weapons (such as fire-and-forget missiles) and defensive weapons (such as surface-to-air missiles) alike for decades already. Regardless of these developments, however, there is as of yet no internationally agreed definition of an autonomous weapon. This is because any workable definition will have to make a clear distinction between existing weapons with autonomous functions and future autonomous weapons. For the purpose of this research report, an autonomous weapon was defined as a weapon that can select and engage targets without further intervention by a human operator. Indeed, all the current definitions adopted by governments, experts and non-governmental organizations alike capture this characteristic nature of autonomous weapons systems: a weapon is only autonomous if the critical functions for using potentially lethal force are performed autonomously, keeping humans out of the loop. In this, the term 'loop' refers to the decision-making process for selecting and attacking targets. This may cover only the critical processes (target selection and engagement) carried out autonomously by the weapon or the entire targeting process in which humans play a decisive role. From the perspective of international humanitarian law, it is advisable that the term 'loop' be interpreted in the latter, wider sense to cover also the various processes preceding the actual selection of and attack on a specific target, including such tasks as formulating objectives, target selection, weapon selection and implementation planning – processes that must also take account of the potential consequences for civilian populations.

Owing to the considerable advantages that autonomous systems offer through computers collecting and processing data faster than humans, thereby facilitating more effective defence against incoming missiles, partially replacing humans on the battlefield, especially in environments where humans cannot survive, thereby reducing the risk to friendly troops, and potentially also helping to limit the number of casualties among civilians, it is most likely indeed that autonomous weapons systems will in the foreseeable future be developed and deployed to attack specific types of targets and carry out defensive tasks. Thus, if armed forces are to remain technologically

advanced, autonomous weapons will have a role to play. And yet, the deployment of such weapons must always involve meaningful human control.

Discussing 'meaningful human control' in the context of fully autonomous weapons systems may seem paradoxical – after all, if there is any meaningful human control, then there cannot be full autonomy. In an actual fact, the notion of 'meaningful human control' can be seen to amount to a ban on full autonomy over certain critical functions of a weapons system, for it becomes significant at various stages of the targeting process. This is because there are various moments at which humans make decisions concerning the use of force and the relevant parameters, whether in adopting rules of engagement, deciding to deploy an autonomous weapon or programming target categories. This in turn means that several people can usually be held accountable for those decisions. For example, as noted above, a commander who decides to deploy an autonomous weapon and the operator who activates it can both be held accountable. Commanders can also be held accountable for humanitarian law violations of their subordinates, if it can be demonstrated that they failed to properly supervise them. The various levels of decisions that go into a targeting process can be seen to reduce the risks associated with the deployment of an autonomous weapon, in particular those concerning the weapon's specific autonomous tasks, its operational environment, the duration of its deployment and its geographical range and mobility.[233] The more limited a weapon's tasks, the less dynamic its operational environment, the shorter the duration of its deployment and the more restricted its mobility, the more predictable the effects of its deployment are likely to be. In such situations, human control is more pronounced. In contrast, a more complex weapon that travels through rapidly changing environments for longer periods of time carries with it a greater risk of unexpected or unpredictable outcomes. In such situations, it is harder to exercise human control. And yet, even in such situations, after an attack has been carried out, it is important to establish

233 See Neil Davison, *Characteristics of Autonomous Weapon Systems*, Presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons 14 April 2015, Geneva, Switzerland, available at <www.unog.ch/80256EDD006B8954/(httpAssets)/37D5012BBF52C7BBC1257E2700599465/$file/Characteristics+of+AWS+ICRC+speaking+points+14+Apr+2015.pdf>.

its effect. Has the objective been achieved? This is the final stage of the targeting process and any feedback thus provided to a command centre where people track the performance of autonomous weapons is instrumental to strengthening meaningful human control over AWS.

As long as the decision to deploy autonomous weapons remains with humans, there is no accountability gap in international humanitarian law for AWS. For this, the existing legal regime is an adequate formal legal framework for holding offenders accountable. At any rate, there is no reason to assume that there will be any erosion of the liability under criminal law of commanders, subordinates or those in positions of political or administrative responsibility during the next decade. They are responsible for deciding whether deploying and activating autonomous weapons in a given context is consistent with the requirements of international humanitarian law and ethically justified. Likewise, there are no gaps in state responsibility as regards the deployment of autonomous weapons. However, compared to the deployment of weapons that require continuous human operation, such as those employed by a rifleman or by a fighter pilot during aerial combat, there is a shift in accountability in the case of autonomous weapons. This is because the deployment of autonomous weapons does not involve a decision to attack a specific target; rather, that decision is implicit in the decisions to deploy and activate them. As a result, accountability lies primarily with the commander who decides to deploy the weapon and the soldier who activates it, as opposed to a soldier who selects and attacks specific targets. This means that commanders and soldiers who are involved in the deployment of autonomous weapons must be well trained and well informed as regards their potential effects. They are required to make judicious decisions concerning distinction, proportionality and precaution without knowing which specific targets will be attacked. That is to say, there has to be meaningful human control throughout the targeting process.

The basic norms of international humanitarian law strictly regulate the deployment of autonomous weapons too. Any deployment that does not comply with these norms is unlawful. As a result, commanders can be held accountable for reckless deployment of autonomous weapons that results in violations of international humanitarian law. Factors such as the interval between the weapon's activation (which is the last moment at which distinction, proportionality and precaution can be considered) and

the actual attack on a target, as well as the complex nature of autonomous weapons, give rise to a need for greater restraint in their deployment. These factors cannot be invoked to evade accountability by arguing that certain consequences were unforeseeable, however.

There seems to be an emerging international consensus on the usefulness of the concept of 'meaningful human control'. Although there is, as of yet, no general agreement on its definition, it is widely acknowledged that the concept can serve as a criterion for distinguishing between acceptable and unacceptable types of autonomous weapons and the deployment of AWS. It also seems sufficiently open-ended to encapsulate even the ethical concerns of the more restrictive – and thus less palatable – notion of 'human dignity' while at the same time allowing sufficient room for striking a balance between prohibition and permission of autonomy.[234] That there is potential for compromise can be seen, for example, in the final statement issued by Germany in the 2015 CCW conference on autonomous weapons systems, announcing that it 'will not accept that the decision over life and death is taken solely by an autonomous system without any possibility for a human intervention'.[235] As has been noted by commentators, this statement 'leaves significant space for requiring different levels of control and for demarcating critical functions that would require high levels of human control from less critical functions that would require lower or no direct human control'.[236]

* * *

The main challenge now resides in the determination of relevant degrees and modes of 'meaningful human control' over autonomous weapons sys-

234 This is a point made by Nehal Bhuta, Susanne Beck and Robin Geiß in 'Present Futures: concluding reflections and open questions on autonomous weapons systems' in Nehal Bhuta et al. eds, *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 347–383 at 382.

235 Final Statement by Germany, CCW Expert Meeting on Lethal Autonomous Weapons Systems, 13–17 April 2015, Geneva, available at <www.unog.ch/80256EDD006B8954/(httpAssets)/07006B8A11B9E932C1257E2D002B6D00/$file/2015_LAWS_MX_Germany_WA.pdf>.

236 Bhuta, Beck & Geiß, 'Present Futures', *supra* note 185 at 382.

tems as well as in the identification of those critical functions and parts of the targeting process over which meaningful human control must remain. To this effect, participants in the CCW meetings should as soon as possible reach for an agreement that the basic assumption underlying all future regulatory approaches is the requirement that decisions concerning critically important functions and thus critically important legal interests – such as the right to life and the right to bodily integrity – may not be delegated to fully autonomous systems. Decisions on life and death must always be under 'meaningful human control' and thus be subject to the ultimate decision of a human being. Member states of the European Union should also seek to coordinate their positions on this issue.

Governments should also advocate (in the upcoming CCW meetings and elsewhere) a more widespread implementation of the procedures relating to Article 36 of First Additional Protocol to the Geneva Conventions at national level, greater transparency concerning the outcomes of these procedures and – crucially – more international information sharing. The Article 36 procedure is to be strictly applied in procuring autonomous weapons, for which the concept of meaningful human control should serve as a benchmark.

In light of the importance of attributing responsibility and accountability, governments should in procuring autonomous weapons ensure that the concept of morally responsible engineering is applied during the design stage and that weapons are extensively tested under realistic conditions. They are also to ensure that ethics training programmes for military personnel – in particular commanders – devote attention to ethical issues relating to the deployment of autonomous weapons.

Governments should at the international level (especially within the CCW framework) promote an international process that will lead to the formulation of either a framework convention or, at the minimum, an interpretative guide that clarifies the current legal landscape with regard to the deployment of autonomous weapons. Such a document would, inter alia, list best practices on such issues as the role of meaningful human control in the Article 36 procedure and in relation to the deployment of autonomous weapons.

More generally, governments should remain actively involved in discussions within the CCW framework on the legal, ethical and policy imp-

lications of developments in the field of autonomous weapon systems. In this regard, a special emphasis should be placed on the importance of conducting a public debate on new technologies. Governments should thus maintain close contacts with non-governmental organizations, the scientific community and other interested parties regarding this issue.