Journal of Conflict & Security Law © Oxford University Press 2013; all rights reserved.
For permissions, please e-mail: journals.permissions@oup.com
doi:10.1093/jcsl/krt002
Advance Access published on 20 February 2013

The Challenge of Autonomous Lethal Robotics to International Humanitarian Law

Chantal Grut*

Abstract

The concept of a truly autonomous weapons system—a system which is capable of operating itself, independently from human oversight—sounds more like science fiction than science fact. However, the reality is that weapons development is increasingly moving in this direction. Despite reassurances that humans will always be 'in the loop', significant amounts of autonomy have been given to certain weapons systems already. Such weapons present unique regulatory problems, arising not so much from their nature as *weapons*, but from their replacement of the human role in war and killing. This article considers the implications of increasing weapon autonomy for the humanitarian law principles of distinction and proportionality, and the concept of accountability for breaches of international humanitarian law.

1. Introduction

As weapons technology becomes more and more advanced, humans are moving further and further away from the battlefield. We already live in a world of robotic warfare, in which a pilot sitting in an operating room in the USA can control an unmanned aerial vehicle or 'drone' to conduct lethal targeting operations on the other side of the world. In a sense, weapons development has always been moving in this direction, with the goal of removing human personnel as far from the risk of harm as possible.[1] The next step may remove the human from the process altogether.

This step is towards autonomous weapons systems. Currently most military robots, like drones, are controlled by a human operator, but in the future, it seems likely that some such robots will control themselves. Autonomy can come in varying degrees. For example, even simple weapons systems like landmines have a very low level of autonomy, in the sense that once laid, they are not

* LLM (Columbia) LLB(Hons) BA (Well). International Legal Consultant, Rosenstock Legal Services, Kabul, Afghanistan.

1 P Asaro, 'How Just Could a Robot War Be?' in P Brey, A Biggle and K Waelbers (eds), *Current Issues in Computing and Philosophy* (IOS Press 2008) 50, 57. For the same point but from a different perspective, see: C Carpenter and L Shaikhouni, 'Don't Fear the Reaper: Four Misconceptions About How We Think About Drones' *Foreign Policy* (7 June 2011).

controlled by the party that laid them—they will automatically explode when triggered.[2] Other already existing weapons systems have a higher level of autonomy. For example, the PHALANX system used by the US Navy is able to autonomously detect, track, and fire at anti-ship missiles.[3] Autonomous sentry guns are set up at the border between North and South Korea that will automatically fire at objects within range.[4] Weapons systems currently being designed and developed are likely to 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 human intervention. According to some commentators, weapons systems incorporating strong artificial intelligence may even be capable of learning and making their own decisions about *how* they decide to conduct targeting.[5] While a number of automated weapons systems with varying degrees of human oversight are already in use (including systems which can make the 'decision' to kill without human input), 'truly autonomous' weapons—which would incorporate strong artificial intelligence and would be capable of choosing their own course of action to reach a desired goal - have not yet been developed.

Yoram Dinstein in a recent edition of this journal declined to address the problems presented by autonomous weapons on the basis that 'Obviously, AI robots do not loom on the proximate horizon, and there does not seem to be any particular urgency in resolving the consequences of their putative action'.[6] But while autonomous killer robots might sound like science fiction, they are much closer to reality than is often appreciated. See for example the drones capable of navigating their own flight paths, developed or being developed by Northrop Gunman, Carnegie Mellon and MIT amongst others;[7] the commissioning of

[2] J Kellenberger, *International Humanitarian Law and New Weapon Technologies,* (Keynote Address, 34th Round Table on Current Issues of International Humanitarian Law, San Remo 8–10 September 2011) <http://www.icrc.org/eng/resources/documents/statement/new-weapon-technologies-statement-2011-09-08.htm> (accessed 4 February 2013).

[3] See: United States Navy Fact File, 'MK 15 – Phalanx Close-In Weapons System (CIWS)' (24 April 2012, 5:30 PM) <http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2> (accessed 4 February 2013); A M Drake, 'Current U.S. Air Force Drone Operations and Their Conduct in Compliance with International Humanitarian Law – An Overview' (2010–2011) 39 Denv J Intl L & Poly 629, 652.

[4] See: Asaro (n 1) fn 2; B Gogarty and M Hagger, 'The Laws of Man over Vehicles Unmanned: The Legal Response to Robotic Revolution on Sea, Land and Air' (2008) 19 J L Inf & Sci 73, 139.

[5] See: Kellenberger (n 2); R Sparrow, 'Killer Robots' (2007) 24 J Applied Phil 62, 64–65.

[6] Y Dinstein, 'The Principle of Distinction and Cyber War in International Armed Conflict' (2012) 17 JCSL 261, 268.

[7] S Ackerman, 'The Navy's Unmanned, Autonomous "UFO"' (Wired 31 July 2012) <http://www.wired.com/dangerroom/2012/07/x47b/#more-88006> (accessed 4 February 2013); J Paur, 'MIT Robot Plan Puts Autonomy Back Into Autonomous Flight' (Wired, 17 August 2012) <http://www.wired.com/autopia/2012/08/mit-autonomous-robot-plane/> (accessed 4 February 2013); O Koski, 'In a First,

research by the US Air Force into software which would allow drones to land and take-off autonomously;[8] and 'swarming' drones, which by definition operate and move together autonomously.[9]

The US Air Force in its 'Unmanned Aircraft Systems Flight Plan 2009–2047' discusses the movement towards and importance of increased autonomy, which is stated to be necessary to meet 'extreme performance parameters'.[10] Full autonomy is recognized as the long-term goal. The US Air Force also recognizes the importance of grappling with the legal and ethical implications of these developments, stating that 'Authorizing a machine to make lethal combat decisions is contingent upon political and military leaders resolving legal and ethical questions'.[11] Britain's Ministry of Defence has also noted the movement towards greater autonomy, and the fact that automated (if not truly autonomous) weapons are already being fielded.[12] As Peter W Singer points out, robots need to get higher on the autonomy scale in order to produce any cost or manpower savings above conventional weapons systems. At the same time, increased autonomy is becoming cheaper and more accessible, and despite official assurances of the importance of keeping a human 'on the loop', weapons development is rapidly moving humans further and further out of the equation.[13]

Some advocates have suggested that autonomous robots might be able to perform in compliance with international humanitarian law (IHL) better than

Full-Sized Robo-Copter Flies With No Human Help' (Wired, 14 July 2010) <http://www.wired.com/dangerroom/2010/07/in-a-first-full-sized-robo-copter-flies-with-no-human-help/> (accessed 4 February 2013).

8 S Ackerman, 'This Man Is Not An Agent of Self-Aware Killer Drones' (Wired, 22 April 2011) http://www.wired.com/dangerroom/2011/04/this-man-is-not-an-agent-of-self-aware-killer-drones/. See also: Defense Systems Staff, Navy, 'Lockheed partner on autonomous controls for VTOL drones' (Defense Systems, 10 October 2012) <http://defensesystems.com/articles/2012/10/10/navy-lockheed-autonomous-uav--controls.aspx?admgarea=DS> (accessed 4 February 2013).

9 'Boeing Technology Allows Drones to Swarm Like Insects' *Chicago Tribune* (Chicago, 9 August 2012) <http://articles.chicagotribune.com/2012-08-09/business/chi-boeing-technology-allows-drones-to-swarm-like-insects-20120809_1_drones-swarm-scaneagle> (accessed 4 February 2013). See also J Emspak, 'Robo-Bee to get Brain for Autonomous Flight' *Discovery News* (2 October 2012) <http://news.discovery.com/tech/robo-bees-to-get-green-brain-121002.html> (accessed 4 February 2013).

10 United States Air Force, 'Unmanned Aircraft Systems Flight Plan 2009–204,' 30 (Washington DC, 18 May 2009).

11 ibid 30. See also H Roff, 'When U.S. Weapons are Autonomous, Who is Responsible?' *Huffington Post* (Canada, 27 September 2012) <http://www.huffingtonpost.ca/heather-roff/the-dods-new-moral-code-f_b_1910608.html> (accessed 4 February 2013).

12 Ministry of Defence (UK), 'Joint Doctrine Note 2/11 The UK Approach to Unmanned Aircraft Systems' (30 March 2011) para 504.

13 PW Singer, *Wired For War: The Robotics Revolution and 21st Century Conflict* (The Penguin Press, 2009) 75, 123.

humans can.[14] While the claim that robots could ever adequately make the kinds of highly contextual analyses that IHL requires is controversial and uncertain,[15] in other aspects of lethal targeting robots already exceed the capabilities of humans, for example in the speed with which they can respond to threats.[16] Increasingly, military systems are becoming 'too fast, too small, too numerous and will create environments too complex for humans to direct'.[17] The increasing ability of robots to respond to threats in ways humans are not capable of controlling indicates that the movement towards greater autonomy may be inevitable.[18] This movement is also a natural consequence of the increasing use of remotely controlled robots such as drones, for which a significant weakness is the link between the operator and the remote platform which can be easily lost as a result of technical mishaps or hacking.[19] When a human operator loses contact with a drone, there is a concern that the expensive system should be able to independently protect itself from attack, rather than simply giving itself up to enemy forces, or automatically returning to base and potentially leading the enemy back with it.[20]

The rapid advancement of robotic technologies, and autonomous technologies in particular, presents some challenges to IHL which the international community needs to start considering. This article will address two key concerns presented by autonomous weapons systems: their implications for the principles of distinction and proportionality; and the challenge they present to accountability and enforcement. This article will then discuss the possibilities for autonomous weapons regulation.

[14] See: G Marchant and others, 'International Governance of Autonomous Military Robots' (2011) XII Colum Sci & Tech L Rev 272, 279–80. See generally: RC Arkin, *Governing Lethal Behavior in Autonomous Robots* (Chapman-Hall, 2009).

[15] See for example: Asaro (n 1) 61; DM Stewart, 'New Technology and the Law of Armed Conflict: Technological Meterorites and Legal Dinosaurs?' (2011) 87 Intl Stud Ser US Naval War Col 271, 282; JF Murphy, 'Mission Impossible? International Law and the Changing Character of War' (2011) 87 Intl Stud Ser US Naval War Col 13, 27. See also: ICRC Report of the 31st International Conference of the Red Cross and Red Crescent, 'International Humanitarian Law and the Challenges of Contemporary Armed Conflicts' (28 November 28–1 December 2011) 40; J McClelland, 'The Review of Weapons in Accordance with Article 3 of Additional Protocol I' (2003) 85 Int Rev Red Cross 397, 408.

[16] See: General Norton Schwartz, Chief of Staff, Air Force Association Convention Speech (15 September 2009) <http://www.af.mil/shared/media/document/AFD-090915-240.pdf> (accessed 4 February 2013); United States Air Force, 'Unmanned Aircraft Systems Flight Plan 2009–204' (Washington DC, 18 May 2009) 33; 41.

[17] TK Adams, 'Future Warfare and the Decline of Human Decisionmaking' (2001–2002) Parameters 58, Vol. XXXI.

[18] Marchant and others (n 14) 276; Arkin (n 14) 7–10; Sparrow (n 5) 64; G Bekey, *Autonomous Robots: From Biological Inspiration to Implementation and Control* (MIT Press, 2005).

[19] See: Drake (n 3) 647; Gogarty and Hagger (n 4) 138.

[20] Singer, *Wired For War* (n 13) 127.

2. Autonomous Robotics and the Law of Armed Conflict

A state's choice of methods and means of warfare is not unlimited.[21] A weapon may present problems for or be contrary to IHL because of the manner in which it is used; or more rarely, the weapon itself may be inherently problematic. International agreements exist to specifically regulate a number of inherently problematic weapons, such as landmines,[22] poisonous gases,[23] blinding lasers,[24] and expanding bullets,[25] but no such regime exists for autonomous weapons.

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 international law. While a number of states, including some not party to Additional Protocol I, have adopted such mechanisms in theory,[26] Article 36 does not seem to have been particularly effective in flagging concerns about autonomous weapons.[27] Few states actively engage with this kind of weapons review process, and such review does not seem to have led to the prohibition of any development, acquisition, or adoption of new weapons.[28] For these reasons, it would be difficult to argue that Article 36 itself is part of customary international law, although the International Committee of the Red Cross (ICRC) has made the argument that the systematic assessment of the legality of all new weapons is an obligation on all states, flowing logically

[21] 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 UNTS 3 (Protocol I) art 35(1). See also: Convention (IV) respecting the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land (The Hague, 18 October 1907) art 22.

[22] Convention on the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction ('Ottawa Treaty') (opened for signature 18 September 1997, entered into force 1 March 1999) 2056 UNTS 211; Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on 3 May 1996 (Protocol II to the 1980 Convention on Certain Conventional Weapons as amended on 3 May 1996).

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

[24] Protocol on Blinding Laser Weapons (Protocol IV to the 1980 Convention on Certain Conventional Weapons), 13 October 1995.

[25] Declaration (IV,3) concerning Expanding Bullets, the Hague, 29 July 1899.

[26] ICRC Study on Customary International Humanitarian Law, Rule 71 <http://www.icrc.org/customary-ihl/eng/docs/v1_cha_chapter20_rule71> (accessed 4 February 2013).

[27] See discussion below: 'Looking Ahead'.

[28] Internationally, the only weapons system to have been regulated before being significantly deployed to ill effect is blinding lasers.

from the prohibition on illegal weapons and the restrictions on means and methods of warfare.[29]

Given the lack of detail in Article 36, and the dearth of state practice from which to draw, the scope of the obligation, and how compliance with it would look in practice, is not at all clear. However, it is clear that the provision does not only concern inherently unlawful weapons, but also covers those, like many autonomous weapons, which have the capacity for great precision but may also be used in a way that is not consistent with IHL. In regards to automation in particular, the commentary to Protocol I notes that: '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.'[30] The ICRC has also suggested that any legal review should include consideration of the technical performance of the weapon.[31]

Two bedrock principles of IHL are the principle of distinction and the principle of proportionality. The principle of distinction requires that parties to armed conflict distinguish between civilian populations and objects, and military objectives. Only military objectives may be targeted in an attack.[32] Weapons which are by nature indiscriminate (in all or in some circumstances) are prohibited, in international[33] and non-international armed conflicts.[34]

The principle of proportionality prohibits an attack if the incidental civilian harm is excessive in relation to the concrete and direct military advantage anticipated by the attack.[35] Parties to armed conflict are required to take all

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

[30] 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) para 1476.

[31] International Committee of the Red Cross (n 29) para 1.3.2.

[32] Protocol I additional to the Geneva Conventions (n 21) arts 48, 51, especially 51(4). See also: ICRC Study on Customary International Humanitarian Law, Rule 7 <http://www.icrc.org/customary-ihl/eng/docs/v1_cha_chapter2_rule7> (accessed 4 February 2013).

[33] Protocol I (n 21) art 51(4); Rome Statute of the International Criminal Court (opened for signature 17 July 1998, entered into force 1 July 1992) 2187 UNTS 90 art 8(2)(b)(xx); Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion) [1996] ICJ Reports 226 para 78 (July 8); ICRC Study on Customary International Humanitarian Law (n 26) Rule 71.

[34] ibid. See also: JM Conde Jimianián, 'The Principle of Distinction in Virtual War: Restraints and Precautionary Measures under International Humanitarian Law' (2010–2011) 15 Tilberg L Rev 69, 75.

[35] Protocol I (n 21) art 51(5)(b); ICRC Study on Customary International Humanitarian Law (n 26) Rule 14.

feasible precautions to ensure that targets are in fact military objectives; to avoid civilian harm; and to ensure compliance with the principle of proportionality.[36]

3. Distinction, Proportionality and Autonomy

Some commentators have argued that, in the future, robots may be more adept at complying with IHL rules and principles like distinction and proportionality than humans are. Gary Marchant and others argue 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.[37] While in theory this is a logical argument, it is questionable how realistic it 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.[38]

Another perceived advantage (or flaw, depending on your perspective) of robots is their lack of emotion. Unlike humans, robots can be designed without emotions to cloud their judgment, such as fear, anger, hysteria, or frustration. There would supposedly be no atrocities like the My Lai massacre in an autonomous robot war.[39]

On the other hand, the kind of analysis that is generally required by the principle of distinction and the principle of proportionality is highly complex and highly contextual—the kind of analysis that the human mind is uniquely adept at. It is debatable whether robots will ever have the same level of ability to distinguish civilian objects from legitimate military targets.[40] 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 in

[36] ibid art 57; ICRC Study on Customary International Humanitarian Law (n 26) ch 5.

[37] Marchant and others (n 14) 279.

[38] Singer, *Wired for War* (n 13) 127; PW Singer, 'The Ethics of Killer Applications: Why is It so Hard to Talk about Morality when it comes to New Military Technology' (2010) 9 J Mil Ethics 299, 303–04.

[39] See: Marchant and others (n 14) 280; ICRC Report of the 31st International Conference of the Red Cross and Red Crescent (n 15) 40; Asaro (n 1) 61; Singer, *Wired for War* (n 13) 63.

[40] RM O'Meara, 'Contemporary Governance Architecture Regarding Robotics Technologies: An Assessment' in P Lin, K Abney and GA Bekey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press, 2012) 129, 165; McClelland (n 15) 408; ICRC Report of the 31st International Conference of the Red Cross and Red Crescent (n 15) 40.
For the argument that robotic systems will never have the kind of capabilities needed, see: N Sharkey, 'Cassandra or False Prophet of Doom: AI Robots and War' (2008) 23(4) IEEE Intelligent Systems 14, 16–17; N Sharkey, 'The Ethical Frontiers of Robotics' (2008) 32 Science 1800, 1800–01.

which we cannot even agree on the exact circumstances under which it is legal to lethally target civilians involved to varying degrees in an armed conflict.[41]

One suggestion to avoid this problem in the context of distinction analysis is to have robots only target other weapons systems, not humans.[42] For example, if an autonomous drone is only programmed to fire at tanks, then the scope for mistakenly targeting civilian objects is significantly reduced (although not eliminated—it is unclear for example how a robot would recognize a tank that was a museum exhibit, or a tank that had been abandoned in a civilian area). This is one of the reasons why existing autonomous systems like the PHALANX are not so concerning—it is designed only to shoot down missiles heading towards a ship, and the civilian objects that could be mistaken for an incoming missile, on the ocean, are limited (but do exist, as the case of the *USS Vincennes*, discussed below, illustrates). 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 defense 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 autonomous weapons such as drones, which are mobile and increasingly capable of directing their own flight paths; as compared to automatic sentry guns which could, for example, be placed only in well-marked areas where the presence of civilians is highly unlikely.

How an autonomous system could undertake a proportionality analysis is even more fraught. To start off with, there is no clear metric to what the principle of proportionality requires. 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 combatant killed; or two civilian casualties per unit commander; or three civilian casualties per tank destroyed. 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.

These are all highly contextual, and inevitably somewhat subjective questions. 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,

[41] For an overview of this fraught debate, see: Human Rights Institute, Columbia Law School, 'Targeting Operations with Drone Technology: Humanitarian Law Implications' Background Note for the American Society of International Law Annual Meeting, 25 March 2011.

[42] M Wagner, 'Taking Humans Out of the Loop: Implications for International Humanitarian Law' (2011) J L Info & Sci at fn 29; JS Canning, 'A Concept for the Operation of Armed Autonomous Systems on the Battlefield', 3rd Annual Disruptive Technology Conference (2006).

making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack'.[43] Another factor in the equation is the extent to which armed forces (or robots themselves) are required to put themselves at risk in order to avert civilian casualties on the other side—an element of the requirement to take precautions[44] in attacks and to 'do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects'. The extent to which armed forces (or their robots) are required to take risks upon themselves in order to avoid civilian casualties is a fraught question on which there is little agreement.[45]

Given that the principle of proportionality cannot be reduced to any kind of formula, and that we struggle to articulate the edges of the prohibition, it is difficult to imagine how one could even begin to try and write the prohibition into a piece of software coding.[46] Even if we could formulate a rule that was capable of being written into code, we would then come back to some of the same difficulties faced in the context of distinction analysis—it is unclear how a robot 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. Given the current level of technological development, it would be pre-emptive to say that robots will never be able to make these kinds of determinations satisfactorily, although at this point in time the barriers might look insurmountable.

These issues are somewhat unique in the weapons development context. Weapons are usually criticized as indiscriminate because of a lack of sophistication and precision, as is the case with landmines, cluster bombs, and certain kinds of rockets;[47] or because of the breadth of their lethal effect, as is the case with biological and nuclear weapons.[48] The concerns presented by autonomous weapons are more subtle. They are at the other end of the scale of sophistication from landmines and cluster bomb fragments, which, once deployed, have no ability to distinguish between targets and will harm whoever is unfortunate enough to set them off. To the contrary, autonomous weapons are being designed as extremely sophisticated systems which are capable of undertaking

[43] *Prosecutor v Galic* (Judgment) ICTY-98-29 (5 December 2003) 43 ILM 794 at 808.

[44] See discussion above. Protocol I (n 21) 57.

[45] See: Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia (2000) 39 ILM 1257, 1271. In the longer term, there is also a concern that as robotic and autonomous weapons became a larger portion of a nation's arsenal, that the ability to undertake on the ground, human investigations in the future might be limited, and the standard for 'everything feasible' referred to in Article 57 of Protocol I could be more limited. See for example: Asaro (n 1) 59.

[46] Wagner (n 42) 8–9.

[47] ICRC Study on Customary International Humanitarian Law (n 26) Rule 71, at fns 40, 42, 43. See also, for example: Cluster Munition Coalition (20 April 2012, 10:00 AM) <http://www.stopclustermunitions.org/> (accessed 4 February 2013).

[48] ICRC Study on Customary International Humanitarian Law (n 26) Rule 71, at fns 33–34.

distinction, but which may simply not do the job as well as human-controlled weapons can. The role they play as a 'weapon' is not as concerning as the role that they play as their own operator. Similarly there is no concern around the nature of the physical effect of the weapon, which may also be designed to be very limited and precise, at the opposite end of the destructive scale from the nuclear weapon, for example.

4. Accountability

Autonomous weapons systems challenge the notions of human accountability which lie at the heart of modern IHL and the system of international criminal justice that has developed to support and enforce it. This presents a serious challenge to the normative framework of IHL,[49] and at worst, might incentivize the creation of weapons systems specifically so that the state and individuals can avoid liability for the conduct of war.[50] If autonomous robots are replacing the role of humans in warfare, then where will accountability lie when something goes wrong? Should we be concerned about the development of autonomous systems if reliance on them could result in a situation where nobody can be held accountable?

A common response to the prospect of full robotic autonomy, without human control or responsibility, is denial. In the words of Peter W Singer, the response is always, '"No, no, no. We'll always have a man in the loop." . . . And yet . . . we have systems right now that we're already granting massive amounts of autonomy to.'[51] Even if we do keep a human 'in the loop', it is not necessarily clear that this will always provide the level of human control, oversight, and responsibility that is appropriate. Further, the more truly autonomous a robot is, the more unrealistic and perhaps unfair it is to expect a human to be responsible for its actions.

Even with relatively low-tech semi-autonomous systems, human operators can be suspect to what is known as 'automation bias'.[52] A tragic example is the shooting down of Iran Air Flight 655 by the Navy ship *USS Vincennes* in 1988. The *USS Vincennes* utilized a computer system known as Aegis, designed

[49] V Kanwar, 'Review Essay: Post-Human Humanitarian Law: The Law of War in the Age of Robotic Weapons' (2011) 2 Harv Natl Sec J 616, 620.

[50] SE White, 'Brave New World: Neurowarfare and the Limits of International Humanitarian Law' (2008) 41 Cornell Intl L J 177, 205–06.

[51] V Wang, 'Robots at War: Popular Science talks to the author of Wired for War: The Robotics Revolutions and Conflict in the 21st Century' Popular Science, 7 April 2009 <http://www.popsci.com/military-aviation-amp-space/article/2009-04/robots-war> (accessed 4 February 2013). See also: Singer, *Wired for War* (n 13) 124; Sparrow (n 5) 69.

[52] ML Cummings, 'Creating Moral Buffers in Weapon Control Interface Design' (2004) 23(3) IEEE Tech & Soc 28, 41; ML Cummings, 'Automation and Accountability in Decision Support System Interface Design' (2006) 32(1) J Tech Stud 23; P Asaro, 'Modeling the Moral User' (2009) 28(1) IEEE Tech & Soc 20, 22.

to defend ships against airborne attacks. Aegis could be set to a number of different modes, which had varying degrees of automation, but in every mode, the human operators had the ability to override the computer. On 3 July 1988, Aegis' radars detected Iran Air Flight 655. The course, speed, radar broadcast, and radio signal coming from the plane all indicated that it was a commercial civilian flight. Aegis, which was in semi-automatic mode, registered 655 as an F-14 Fighter, a plane half the size. While all of the evidence should have made it clear that 655 was a civilian plane, not one of the people on board the *USS Vincennes* were willing to challenge the computer's wisdom. They trusted it, and they authorized it to shoot. Afterwards, the crew realized that they had shot down a civilian plane, killing all 290 passengers including 66 children, in one of the worst aviation incidents in history.[53] As Singer notes, 'The automated Aegis system . . . had been designed for managing battles against attacking Soviet bombers in the open North Atlantic, not for dealing with civilian filled skies in the crowded Gulf.'[54]

The shooting down of Iran Air Flight 655 is a particularly outrageous example of automation bias, because of the wealth of evidence outside of the Aegis system which clearly indicated that the plane was civilian. Other cases are more subtle. For example a very similar event occurred during the 2003 Iraq invasion, when US Patriot missiles shot down two allied planes that had been incorrectly identified as Iraqi rockets:

> There were only a few seconds to make a decision, and so the human controllers trusted the machine on what to fire at. Their role "in the loop" was actually only veto power, and even that was a power they were unwilling to use against the quicker (and what they viewed as better) judgment of a computer.[55]

As 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 becomes even less realistic to expect human operators to exercise significant veto control over their operations. Given what we know about how humans interface with these systems, in many cases it would be unfair to hold an individual operator accountable for failing to hit the 'off' switch. It is only natural, particularly in complex situations, that an operator would be cautious about second guessing the machine's wisdom and running the risk of not preventing a potentially deadly attack. The deficit of human accountability in this scenario is more systemic than individual. If this is the case even for relatively basic, semi-automated systems, then it is even more so the case with advanced and more truly autonomous systems which make their own independent and complex decisions.

[53] Singer, *Wired for War* (n13) 124.
[54] ibid.
[55] ibid.

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 responsible for creating the system in the first place.[56] This is essentially shifting from an IHL framework to a product liability framework, which arguably does not provide a strong enough sanction given the importance of and high stakes involved in IHL compliance. Singer argues:

> While a system may be autonomous, those who created it still hold some responsibility for its actions. Given the larger stakes of war crimes, though, the punishment shouldn't be a lawsuit, but criminal prosecution. If a programmer gets an entire village blown up by mistake, the proper punishment is not a monetary fine that the firm's insurance company will end up paying. Many researchers might balk at this idea and claim that it will stand in their way of their work. But as Bill Joy sensibly notes, especially when the consequences are high, "Scientists and technologists must take clear responsibility for the consequences of their discoveries." Dr. Frankenstein should not get a free pass for his monster's work, just because he has a doctorate.[57]

Generally, however, criminal accountability is predicated on a guilty state of mind, and replacing the responsibility of a solider with the responsibility of a computer technician is not as straightforward as Singer makes it sound. Even assuming criminal liability for manufacturing errors could and should attach, the nature and complexity of these systems would make successfully attaching personal liability to any one programmer or engineer very difficult. Gogarty and Hagger explain how this is the case even in the non-criminal context where there is a lower standard of proof:[58]

> . . . determining fault in complex software and hardware is already difficult. Given that [unmanned vehicles] require systems which are increasingly complex and powerful, the ability of negligence to reach into the maze of complexity and extract a responsible part is likely to be limited. Moreover, it is limited by salient considerations of causal, physical and circumstantial proximity which seek to place a reasonable constraint on unfair or burdensome duties being imposed on whose who are simply too far removed from the act that caused harm. It is unlikely that a court would impose liability on a computer programmer whose small piece of code – possibly designed for much more general purposes

[56] See: Murphy (n 15) 28; Gogarty and Hagger (n 4) 123.
[57] Singer, *Wired for War* (n 13) 412.
[58] See also: White (n 50) 209.

> than being used in a [unmanned vehicle] – caused an unforeseen conflict within a massive code library, resulting in a [unmanned vehicle] that acts in an unpredictable or dangerous way.[59]

While the move towards autonomous and highly advanced weapons systems does justify reassessing how we deal with the liability of actors like engineers and programmers, once again, liability may be more systemic than individual. For example, if a software company sells a piece of coding without properly testing it,[60] and the military purchases and fields it knowing this, under weapons acquisitions procedures authorized by Congress, then who should be responsible? The US military currently and quite regularly puts new weapons systems into production and use before they have been properly tested.[61]

Another floated means of establishing accountability for the use and actions of autonomous systems is the doctrine of command responsibility.[62] The elements of command responsibility are set out in Article 28 of the Rome Statute of the International Criminal Court,[63] which states:

[59] Gogarty and Hagger (n 4) 123.

[60] See for example: J Stein, 'CIA Mum on Lawsuit Alleging Drone Targeting Errors' *Washington Post* (Washington DC, 4 October 2010) <http://voices.washingtonpost.com/spy-talk/2010/10/cia_mum_on_lawsuit_alleging_dr.html> (accessed 4 February 2013).

[61] See for example: United States General Accounting Office, Report to the Chairman, Subcommittee on Tactical Air and Land Forces, Committee on Armed Services, House of Representatives, 'Force Structure: Improved Strategic Planning Can Enhance DOD's Unmanned Aerial Vehicles Efforts', 5–7 (March 2004) available at <http://www.fas.org/irp/gao/gao-04-342.pdf> (accessed 4 February 2013); United States Government Accountability Office, Testimony before the Subcommittee on National Security and Foreign Affairs, Committee on Oversight and Government Reform, House of Representatives, 'Defense Acquisitions: DOD Could Achieve Greater Commonality and Efficiencies among its Unmanned Aircraft Systems', 23 March 2010, 2, <http://www.gao.gov/assets/130/124311.pdf> (accessed 4 February 2013); United States Government Accountability Office, Testimony before the Subcommittee on National Security and Foreign Affairs, Committee on Oversight and Government Reform, House of Representatives, 'Defense Acquisitions: DOD Could Achieve Greater Commonality and Efficiencies among its Unmanned Aircraft Systems' 23 March 2010, 2 <http://www.gao.gov/assets/130/124311.pdf> (accessed 4 February 2013); United States Government Accountability Office, Report to Congressional Committees, 'Defense Acquisitions: Assessments of Selected Weapon Programs', March 2012, GAO-12-400SP <http://www.gao.gov/assets/590/589695.pdf> (accessed 4 February 2013).

[62] See: White (n 50) 205–206; Stewart (n 15) 292; Murphy (n 15) 28.

[63] This is a codification of existing customary international law. See: ICRC Study on Customary International Humanitarian Law (n 26) Rule 153; K Dormann, 'War Crimes under the Rome Statute of the International Criminal Court, with a Special Focus on the Negotiations on the Elements of Crimes' (2003) 7 Max Plank Ybk of UN Law 341, 345.

> 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 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.[64] It has been suggested, however, that the concept could be expanded to cover responsibility for autonomous weapons.[65] The proposal is sensible in that it squarely addresses the unique aspect of autonomous systems which makes them so concerning—that rather than just being another kind of weapon, they are really replacing the human role of the solider.

[64] As well as references to 'forces' which is clearly intended to refer to people, robots cannot really commit 'crimes', as they do not have *mens rea* in any real sense: Murphy (n 15) 27.

[65] White (n 50) 205; Drake (n 3) 659.

In the context of what we know about automation bias, it makes sense to place responsibility for autonomous systems at a relatively high level. As the Director for International Law and Cooperation at the ICRC has commented in this context, someone has to be responsible for turning the system on.[66] Some difficulties in accountability could be avoided if that person were a relatively high level commander.

It is likely that the accuracy and reliability of most autonomous systems will not be able to be measured in the usual ways. For example, the accuracy of rockets and bombs can be measured in terms of a radius, and the legitimacy of using the weapon will then depend, in part, on the number of civilian objects within that radius. The accuracy of an autonomous system might depend, for example, on factors like how many objects it will have to analyze, the ability of its sensors to properly analyze the kinds of objects in range, and whether it is operating in an area where there is greater scope for mistaken targets (such as the crowded skies of the Gulf leading to incorrect determinations of aerial threats). This makes the determination of reliability or accuracy difficult. Such information may also be confidential and not available at every level of the military.[67]

In a sense, this is a round-about way of attaching criminal liability to a failure to comply with the requirement to take precautions in attack.[68] In particular, if a commander or drone operator should have known that an autonomous weapon was likely to commit the equivalent of a war crime but turned it on anyway, then they have not taken 'all feasible precautions in the choice of means and methods of attack with a view to avoiding . . . incidental loss of civilian life'[69] 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'.[70]

However, in order to apply either the doctrine of command responsibility or the duty to take precautions in the context of autonomous weapons systems, the system must have some degree of predictability. Otherwise the commander will rarely be in a position where he or she '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

[66] P 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, San Remo, 8–10 September 2011) <http://www.icrc.org/eng/resources/documents/statement/new-weapon-technologies-statement-2011-09-13.htm> (accessed 4 February 2013).

[67] White (n 50) 206–07.

[68] See: Protocol I additional to the Geneva Conventions (n 21) art 57; ICRC Study on Customary International Humanitarian Law (n 26) Rule 15 <http://www.icrc.org/customary-ihl/eng/docs/v1_rul_rule15> (accessed 4 February 2013).

[69] Protocol I ibid art 57(2)(a)(ii).

[70] ibid.

the system is predictable enough to make even this kind of human accountability realistic or sensible. Marchant and others have noted that it is:

> . . . a common misconception that robots will do only what we have programmed them to do. Unfortunately, such a belief is sorely outdated . . . Now, 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 . . . [71]

Regarding the possibility of autonomous weapons systems with strong artificial intelligence, capable of independent learning, there is also a genuine concern as to whether it will even be possible to program robots so that they will obey orders.[72] While this kind of concern might seem more like science fiction than science fact, it is an implication of the very nature of artificially intelligent and truly autonomous robotics, which is the direction in which weapons development continues to travel.

5. Looking Ahead

In a book on the robotic revolution in warfare, Singer writes:

> In the hundreds of interviews for this book, not one robotics researcher, developer, program manager, or solider using them in the field made a single reference to the ICRC, nor its all-important "four pillars" of international humanitarian law on weapons. That is, not a single organization, research lab, or company working on robotics today is formally linked up with the ICRC or has in place . . . reviews . . . necessary for new weapons. Indeed, the closest one gets to any such legal reviews are some limited efforts by robotic firms to make sure that they don't get sued by customers.[73]

The question of how we should address the issues presented by autonomous weapons systems is not a straightforward one. As robotic capabilities are advancing at a rapid rate, it would pay to pick up the pace of the ethical, legal, and regulatory discussion about their use and development.[74] At least one civil society group, the International Committee for Robot Arms Control (ICRAC) has tried to start this debate. ICRAC advocates for a dialogue

[71] Marchant and others (n 14) 283. See also Sparrow (n 5).
[72] Marchant ibid 285; Sparrow ibid 5.
[73] Singer, *Wired for War* (n 13) 385.
[74] Although for an alternative view, see: Drake (n 3) 659.

including consideration of the potential of military robotics to lower the threshold of armed conflict; prohibiting the development, deployment and use of armed autonomous unmanned systems that can make the decision to kill people; limiting the range and weapons carried by robotic systems in which there is a 'man in the loop'; banning unmanned systems with nuclear weapons and prohibiting the development, deployment and use of robot space weapons.[75]

As a category, autonomous weapons systems could encompass a wide range of potentially quite diverse systems, from fixed sentry guns and missile defense systems to unmanned autonomous drones which direct their own flight, target detection, and target engagement. As such, it is not necessarily the case that every autonomous system should be regulated in the same way, if specifically regulated at all. Ideally new capabilities would be assessed as they were developed (as envisaged by article 36 of Protocol I), but as Singer's above observation illustrates, that is not the existing approach. The history of weapons regulation also suggests that the push needed for regulation usually only comes after the weapon has been put into use and its negative effects made a reality.[76]

Although they are autonomous in a very limited sense, landmines are a kind of indiscriminate autonomous weapon that has been regulated. Although the regulation of more complex autonomous systems is likely to be itself more complex, some lessons can be learned from landmines. Rather than banning landmines outright, Protocol II to the Convention for Certain Conventional Weapons focuses (amongst other things) on where and how landmines are placed and how long they are active for. In the case of landmines, it was their continuing lethal presence long after battles had finished and civilians had moved back home that so amplified their indiscriminate effect. This was mitigated by a requirement that mines either be laid within an area clearly marked as a minefield; or that they are designed so that no more than 10% would fail to self-destruct within 30 days of emplacement, and with an additional self-destruction feature.[77] Using a means or method of laying mines that could not be directed against a military target, or that could be expected to breach the principal of proportionality, is banned.[78] States are required to take all feasible precautions to avoid civilian harm including using alternatives to landmines,[79] and the Protocol made it clear that the state who lays a mine maintains responsibility for it.[80] Despite these limitations on the use and technical specifications of landmines, a large number of states, driven by a civil society movement, came to the conclusion that anti-personnel mines in particular still

[75] International Committee for Robot Arms Control (14 April 2012, 1:14 PM) <http://www.icrac.co.uk/mission.html> (accessed 4 February 2013).
[76] With the possible exception of blinding laser weapons, see: White (n 50) 210.
[77] Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, at Technical Annex 3.
[78] Art 3(8).
[79] Art 3(10).
[80] Art 3(2).

caused too much civilian harm, and so entered into a convention to ban anti-personnel mines altogether.[81]

Similarly to the regulation of landmines, the potentially indiscriminate effect of autonomous weapons which are fixed rather than free-moving (such as sentry guns and anti-missile systems like PHALANX) can be partly controlled by limiting the locations in which they are used and potentially having minimum accuracy or reliability requirements. In large part, the existing use of such systems may already comply with these basic precautions. For example, the use of an automatic sentry gun on the border of North and South Korea may not be so concerning if it is known that no civilians live in the area or have cause to be in the area, and the presence of the sentry guns is clearly signposted. Systems which control their own movement may be more difficult to control, but in some cases it may be possible to only deploy them only in particular locations and contexts in which there is less risk of indiscriminate or disproportionate attack, depending on the particular strengths and weaknesses of the technology.

While there is a lot of skepticism regarding whether robots will ever be capable of adequately complying with IHL while acting autonomously, we do not really know what future developments hold in store in this arena. Given that some developments towards autonomy might result in more discriminatory systems, a blanket ban on autonomy, particularly in the development stage, would not be wise. On the other hand, it is important that as yet unfounded claims about robotic capabilities are not used to justify creating and fielding indiscriminate and unreliable weapons. In light of the level of uncertainty around future capabilities, a prudent course of action would be a moratorium on un-fixed autonomous systems and systems designed to autonomously identify, target and kill human beings. These kinds of highly autonomous systems are on the horizon, and yet the high level artificial intelligence and reliability needed to trust such systems to be sufficiently discriminating is not so advanced.

Whether or not this kind of regulation is realistic in the near future is another question—despite the best efforts of groups like ICRAC, there appears to be little political will to regulate autonomous weapons. This might be partly attributable to public and political disengagement with the issue, which may in part be a result of how these kinds of weapons systems have lowered the moral costs of the wars they are being designed for. Contrary to rudimentary systems like landmines, autonomy is also a cutting edge technology—it is only within the near grasp of some states. So long as some states maintain a significant military advantage by having access to such technology, and as long as such technology is primarily being used against non-state actors,[82] then the international community seems unlikely to find the kind of political will needed to explicitly address these questions in a convention framework. However, even if explicit regulation is not possible, a dialogue on these issues and monitoring of developments would be a start.

[81] Ottawa Convention.
[82] Which is currently the case with lethal manned robots.

6. Conclusion

Autonomous weapons systems present significant challenges to IHL. Their ability to reliably act in a discriminatory and proportionate manner is suspect; and their effect in removing humans from targeting decisions challenges the notions of accountability that the normative and enforcement framework of IHL relies on. Additionally, autonomous weapons systems are likely to present complex *jus ad bellum* issues related to lowering the moral, political and financial cost of warfare for those parties able to develop and deploy them with little human involvement. How to deal with these issues is made all the more complex by the complexity, diversity and uncertainty surrounding autonomous technologies. The international community needs to start considering and debating these issues. Otherwise, the risk is that we delegate too much responsibility to machines before the technology is really capable; and in turn, that we abdicate too much human responsibility for warfare.[83]

[83] See: Singer, *Wired for War* (n 13) 389.