The ethical and legal case against autonomy in weapons systems

Daniele Amoroso[*] and Guglielmo Tamburrini[†]

Abstract

In order to be counted as *autonomous*, a weapons system must perform the critical functions of target selection and engagement without any intervention by human operators. Human rights organizations, as well as a growing number of States, have been arguing for banning weapons systems satisfying this condition – that are usually referred to as autonomous weapons system (AWS) on this account – and for maintaining a meaningful human control (MHC) over any weapons systems. This twofold goal has been pursued by leveraging on ethical and legal arguments, which spell out a variety of deontological or consequentialist reasons. Roughly speaking, deontological arguments support the conclusion that by deploying AWS one is likely or even bound to violate moral and legal obligations of special sorts of agents (military commanders and operators) or moral and legal rights of special sorts of patients (AWS potential victims). Consequentialist arguments substantiate the conclusion that prohibiting AWS is expected to protect peace and security, thereby enhancing collective human welfare, more effectively than the incompatible choice of permitting their use. Contrary to a widespread view, this paper argues that deontological and consequentialist reasons can be coherently combined so as to provide mutually reinforcing ethical and legal reasons for banning AWS. To this end, a confluence model is set forth that enables one to solve potential conflicts between these two approaches, by prioritizing deontological arguments over consequentialist ones. Finally, it is maintained that the proposed confluence model significantly bears on the issue of what it is to exercise genuine MHC on existing and future AWS. Indeed, full autonomy is allowed by the confluence model in the case of some anti-materiel defensive AWS; it is to be curbed instead in the case of both lethal AWS and future AWS which may seriously jeopardize peace and stability.

1. Introduction

A weapons system must perform the critical functions of target selection and engagement without any intervention by human operators in order to be counted as *autonomous*. In the last few years, the development and use of systems satisfying this condition (and usually called AWS – autonomous weapons systems – on this account) have become an issue of common concern for States, International Organizations and NGOs, which have been debating the way forward on AWS, on a fairly regular basis, both in formal and informal fora.[1] Remarkably enough, since 2014, an Informal Meeting of Experts on *Lethal* AWS (a.k.a. as LAWS) has been held annually in Geneva within the institutional framework of the UN Convention on Conventional Weapons (CCW), gathering an ever-increasing number of States. A formal discussion on AWS among CCW State Parties was held at the Fifth Review CCW Conference (12-16 December 2016),[2] where it was decided to establish an open-ended Group of Governmental Experts on Lethal AWS, with the mandate of "submit[ting] a report to the 2017 Meeting of the High Contracting Parties to the Convention consistent with" the recommendations issued in the Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems,[3] namely "to explore and agree on possible recommendations on options related to emerging technologies in the area of L[ethal] AWS, in the context of the objectives and purposes of the Convention, taking into account all proposals – past, present and future".[4]

Awareness about this topic has been raised by the Campaign 'Stop Killer Robots', which was launched in 2013 by an international coalition of NGOs with the primary goal of "ban[ning] lethal robot weapons"[5] – a goal subsequently endorsed, in a 2015 Open letter, by a surprisingly large group of high-profile experts in the fields of robotics and Artificial Intelligence.[6] Over the years, the

---

[1] A fairly comprehensive chronology of the initiatives taken in this respect at international and national levels is available at: www.stopkillerrobots.org.

[2] All transcripts and documents of the CCW Meetings of Experts as well as of the Fifth Review Conference are available at: www.unog.ch.

[3] Final Document of the Fifth Review Conference, 23 December 2016, UN Doc. CCW/CONF.V/10, Decision 1. The Group of Governmental Experts was scheduled to hold its first session from 21 to 25 August 2017 and its second session from 13 to 17 November 2017. Regrettably, however, due to financial constraints, the first session was cancelled. See Stop Killer Robots, *Diplomatic Efforts to Address Killer Robots Falter Convention on Conventional Weapons Talks Cancelled Due to Lack of Funds*, May 30, 2017, http://www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_PR_CCW_30May2017fnl.pdf (accessed June 5, 2017).

[4] Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 10 June 2016, UN Doc. CCW/CONF.V/2, Annex para. 3.

[5] Stop Killer Robots, *Urgent Action Needed to Ban Fully Autonomous Weapons. Non-governmental organizations convene to launch Campaign to Stop Killer Robots*, May 30, 2013, http://stopkillerrobots.org/wp-content/uploads/2013/04/KRC_LaunchStatement_23Apr2013.pdf (accessed June 5, 2017).

[6] Future of Life Institute, *Autonomous Weapons: An Open Letter From AI & Robotics Researchers*, July 28, 2015, https://futureoflife.org/open-letter-autonomous-weapons (accessed June 5, 2017; hereinafter 2015 Open Letter). The Open letter was signed, among others, by Stephen Hawking, Elon Musk, Jaan Tallinn and Frank Wilczek. The call for a ban has been also endorsed by the EU Parliament, which adopted a resolution calling on "the High Representative for Foreign Affairs and Security Policy, the Member States and the Council to: […] (d) ban the development, production

Campaign's focus has apparently shifted from calling a ban on lethal AWS to requesting meaningful human control (MHC) on weapons systems. However, this shift does not represent a radical turn in the overall Campaign strategy. As we shall see, particularly restrictive interpretations of MHC are licensed by some (deontological) arguments against AWS that were formulated by AWS ban supporters even before the MHC issue came to the fore.

One should be careful to note that arguments for banning AWS are not, in general, limited in their scope to *lethal* AWS. One may additionally express concerns about threats for peace and global stability which may arise from AWS which have no direct lethal or sublethal effect on human beings. A significant case in point concerns applications of swarm intelligence technologies which may enable one to level massive attacks on possibly uninhabited enemy infrastructures.[7]

The ambitious goal of aiming for a treaty prohibiting AWS which *in principle* is not limited to *lethal* AWS must address the challenge of merging coherently into a unitary ethical and legal framework both deontological and consequence-looking approaches to AWS, which usually do not share the same background assumptions and may even reach conflicting conclusions. We will address here this substantive merging problem, with the aim of outlining a coherent ethical and legal framework in support of a ban proposal which is not limited to lethal AWS.

One can readily appreciate the existence of a substantive merging problem by reflecting on the variety of assumptions that one finds in both ethical and legal arguments for prohibiting AWS. In normative ethics, reasons for prohibiting AWS have been put forward from both deontological and consequentialist standpoints. Roughly speaking, deontological arguments aim to support the conclusion that by deploying those weapons one is likely or even bound to violate moral obligations of special sorts of agents (military commanders and operators) or moral rights of special sorts of patients (human beings that are potential victims of AWS). Consequentialist arguments do not start from inalienable rights and absolute duties, and rather aim to support the conclusion that AWS which threaten peace and security must be prohibited in order to preserve and enhance collective human welfare. At the same time, however, it is hard to develop sound consequentialist arguments for banning every lethal AWS, including those lethal AWS that fail to pose significant threats to aggregate human welfare. Hence, potential tensions may arise between deontological and

---

and use of fully autonomous weapons which enable strikes to be carried out without human intervention". See European Parliament, *Resolution on the Use of Armed Drones*, 27 February 2014, 2014/2567(RSP), para. 2.

[7] The US Department of Defense is already testing swarms of autonomous micro-drones. See US Department of Defense, *Department of Defense Announces Successful Micro-Drone Demonstration*, January 9, 2017. https://www.defense.gov/News/News-Releases/News-Release-View/Article/1044811/department-of-defense-announces-successful-micro-drone-demonstration/see US Department of Defense 2017 (accessed June 5, 2017).

consequentialist reasons in normative ethics that are advanced from one and the same AWS ban advocacy perspective.

Arguments based on international law have been developing along similar lines. On the one hand, it has been maintained that the deployment of AWS would be in stark contrast with the basic (deontological) tenets of international humanitarian law (IHL), international human rights law (IHRL) and international criminal law (ICL). On the other hand, in an ostensibly consequentialist perspective, the proliferation of AWS has been claimed to make wars easier to wage, with inevitable backlashes on the prohibition on the use of force (art. 2, para. 4 of the U.N. Charter) and, more generally, on the maintenance of international peace and security (art. 1, para. 1 of the U.N. Charter).

The present paper acknowledges some limitations of each pro-ban argument taken in isolation and the potential conflicts that may arise between arguments that are based on deontological and consequentialist assumptions. Nevertheless, it is argued here that these potential conflicts can be defused, thereby enabling one to take advantage for ban advocacy purposes of the complementary and mutually reinforcing character of deontological and consequentialist arguments. This is achieved by introducing a confluence mode, which enables one to solve conflicts between deontological and consequentialist approaches in the AWS domain of application. Moreover, it is maintained that the proposed model is justified by reference to the hierarchy of sources of international legal order. Finally, the implications of this model towards a restrictive interpretation of MHC are explored with reference to a variety of existing and technologically possible AWS. It turns out that this model allows for full autonomy in the case of some anti-materiel defensive AWS, requires one to prohibit the autonomy of lethal AWS in view of deontological reasons, and to curb in view of consequentialist reasons the autonomy of AWS which may have a destabilizing impact, irrespective of whether the latter have any direct lethal effect on human beings.

The paper is organized as follows. In section 2, various requirements that have been put forward for a weapons system to count as autonomous are analyzed. The more liberal requirement on AWS, which is satisfied by various existing weapon systems, is adopted on a precautionary basis for discussion among stakeholders, and compared to more restrictive requirements which end up projecting AWS in some undetermined technological future. Moreover, it is pointed out that the recent shift in focus from the notion of "autonomy" to that of "meaningful human control" (or lack thereof) enable one to sidestep the need to resolve the AWS definitional conundrum in connection with most current debates about AWS (2.1). The merging problem for ethical and legal arguments against AWS is addressed and analyzed by reference to the broad categories of the deontological

and consequentialist approaches in sections 3 and 4, respectively. To combine coherently deontological and consequentialist reasons into a unitary ethical and legal framework, a confluence model is proposed in section 5. This model prioritizes deontological perspectives whenever agent-relative obligations and patient-relative rights are applicable, and shifts by default to agent-neutral consequentialist perspectives whenever those obligations and rights are no longer applicable. In section 6, this confluence model is informatively applied to determine what level of human control is to be exerted on existing weapon systems which, according to the more liberal requirement for autonomy, can already operate in autonomous mode. Section 7 concludes by distilling the legal outputs which can be drawn from the proposed confluence model.

2. On the Notion of Autonomy in AWS Ethical and Legal Debates

Ethical and legal debates on AWS are usually based on a construal of autonomy in terms of task-execution or goal-achievement capabilities and related information-processing skills for technological devices.[8] A demanding requirement along these lines was advanced in a document of the UK Ministry of Defense. There, an autonomous system is required to be "a system which is capable of *understanding higher level intent and direction*. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about the desired state".[9] These requirements are hardly satisfied by any existing weapons system. Moreover, no educated guess can be made about the prospects of constructing machines which are capable of understanding high level intent and direction, unless one redefines these terms operationally and quite differently from common usage. Therefore, on the basis of common linguistic usage, the UK Ministry of Defense definition projects AWS in some undetermined technological future.[10]

The US Department of Defense (DoD) proposed a less demanding requirement on weapons systems to count as autonomous. Indeed, autonomous weapons systems must be capable, "once activated, to select and engage targets without further intervention by a human operator".[11] Much of the ongoing

---

[8] Guglielmo Tamburrini, "On banning autonomous weapon systems: from deontological to wide consequentialist reasons" *in Autonomous Weapons Systems: Law, Ethics, Policy*, Ed. Nehal Bhuta et al. (Cambridge University Press, 2016), 122, 124-127 (hereinafter Tamburrini, *On banning AWS*).

[9] UK Ministry of Defence, *The UK Approach to Unmanned Aircraft Systems*, Joint Doctrine Note 2/11, 30 March 2011, 14 (emphasis added).

[10] Coherently with this premise, the UK has boldly affirmed that AWS "do not, and may never, exist". See the Opening Statement by the UK delegation at the 2016 CCW Meeting of Experts (General Exchange, 1).

[11] US Department of Defense, *Autonomy in Weapons Systems*, Directive 3000.09, 21 November 2012, 13-14 (hereinafter DoD Directive). A largely overlapping requirement for autonomy has been adopted by the International Committee for the Red Cross, which refers to "weapons that can independently select and attack targets, i.e. with autonomy in the 'critical functions' of acquiring, tracking, selecting and attacking targets". See ICRC, *Autonomous weapon systems: Technical, military, legal and humanitarian aspects. Expert meeting, Geneva, Switzerland, 26-28 March 2014*, 1 November 2014, 3.

ethical, legal and political discussion on AWS takes the DoD requirement for autonomy as a starting point, identifying without any further ado an autonomous weapons system with a system satisfying the DoD requirement.[12] Nevertheless, in these discussions the DoD requirement is usually refined and suitably adapted to various classes of AWS, insofar as the required autonomy is always relative to some set of warfare environments, and to the specific perceptual, reasoning and action capabilities that the weapon system is endowed with.[13]

One should carefully note that by taking the DoD requirement as a definition of AWS licenses as autonomous a variety of presently operating weapons systems.[14] These include anti-materiel defensive systems like the German Nächstbereichschutzsystem MANTIS and the Israeli Iron Dome, active protective systems for vehicles (e.g. the South-African/Swedish LEDS-150), loitering weapons systems like the Israeli anti-radiation Harpy and Harop systems, a variety of offensive fire-and-forget munitions, like the British Brimstone missile, and stationary (i.e. non-mobile) robotic sentinels, like the South-Korean Super aEgis II, which surveils the border between North and South Korea, and the gun tower developed by the Turkish company ASELSAN to be installed on the border between Turkey and Syria.[15] Thus, if one accepts the DoD requirement as an adequate criterion to discriminate between autonomous and non-autonomous weapons systems, then one is *ipso facto* acknowledging that some AWS exist and have been actually deployed in various warfare scenarios.[16] Accordingly, the ethical, legal, and political discussions of AWS that are based on the DoD requirement cannot be construed as discussions concerning future combat systems only. This far-reaching implication of the methodological choice of accepting the DoD requirement as a definition in ethical, legal, and political discussions of AWS is aptly illustrated by reference to the

---

[12] See, among many others, Human Rights Watch, *Losing Humanity. The Case against Killer Robots*, November 19, 2012, 7-8 (hereinafter HRW, *Losing Humanity*); Cristof Heyns, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, UN Doc. A/HRC/23/47, 9 April 2013, para. 38 (hereinafter *Heyns Report*); Michael N. Schmitt, "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics." *Harvard National Security Journal Features* 1(2013): 1, 4 (hereinafter Schmitt, *Autonomous Weapon Systems*); Rebecca Crootof, "The Killer Robots Are Here: Legal and Policy Implications." *Cardozo Law Review* 36(2015): 1837, 1847 (hereinafter Crootof, *The Killer Robots Are Here*); Eliav Lieblich and Eyal Benvenisti "The obligation to exercise discretion in warfare: why autonomous weapons systems are unlawful" *in Autonomous Weapons Systems: Law, Ethics, Policy*, Ed. Nehal Bhuta et al. (Cambridge University Press, 2016), 245, 249; Duncan B. Hollis, "Setting the Stage: Autonomous Legal Reasoning in International Humanitarian Law." *Temple International & Comparative Law Journal* 30(2016): 1, 6.

[13] See, for instance, the Working Paper presented by International Committee of the Red Cross (ICRC) at the 2016 CCW Meeting of Experts (Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system, 11 April 2016, 3-4; hereinafter *ICRC Working Paper*).

[14] For a more refined and comprehensive survey on existing autonomous weapons systems, see the dataset processed as part of the project on "Artificial Intelligence, Autonomous Weapons, and Meaningful Human Control" (Principal investigators: Heather Roff and Richard Moyes), available at: https://globalsecurity.asu.edu/robotics-autonomy (accessed June 5, 2017).

[15] Yeni Şafak, *Turkey starts building automatic shooting gun towers at Syrian border*, May 30, 2016, http://www.yenisafak.com/en/news/turkey-starts-building-automatic-shooting-gun-towers-at-syrian-border-2473487 (accessed June 5, 2017). For more information about the ASELSAN Stabilized Gun Platform, see www.aselsan.com.tr.

[16] Crootof, *The Killer Robots Are Here*; *ICRC Working Paper*.

previously mentioned Iron Dome, Brimstone, and Super aEgis II. Let us briefly consider each one of these systems in turn.

The Iron Dome system was primarily designed to counter incoming short-range rockets. Human operators feed the system in advance with information concerning the area to be protected from incoming rockets. On the basis of this information, the system detects the launch of a rocket and tracks its trajectory in order to determine whether the incoming projectile threatens the designated area. Only if a threat is detected, an interceptor missile is fired to destroy the incoming rocket. In principle, human operators can override the system's firing decision.[17] Yet, since veto should be exercised within a prohibitively short time frame (a matter of split seconds), operators will be generally unwilling (and in many cases utterly unable) to challenge the assessment made by the system, which will therefore operate in a *de facto* autonomous mode.[18] Accordingly, the critical target selection and attack functionalities mentioned in the DoD requirement are both performed without a human operator being really able to intervene.

Brimstone is an air-to-ground attack missile which works either as fire-and-forget munition or as laser-guided missile. In the fire-and-forget mode, Brimstone is loaded with targeting data, including data which serve to circumscribe the area within which armored enemy vehicles must be searched for, selected and attacked without any further intervention by human operators. In the laser-guided mode, human operators pick out the specific targets to destroy.[19] Thus, Brimstone missiles satisfy the DoD requirement for autonomy in the fire-and-forget mode, and fail to satisfy the same requirement in the laser-guided mode. The operation mode is selected by human operators on the basis of available information about individual attack scenarios (e.g. considering whether there are civilians or friendly forces in the vicinity of targets).

The DoDAAM system Super aEgis II is a stationary robotic platform designed to replace or to assist South Korean sentinels in the surveillance of the demilitarized zone between North and South Korea.[20] The Super aEgis II can be operated in either unsupervised or supervised modes. In the unsupervised mode, the Super aEgis II identifies and tracks intruders in the demilitarized zone, deciding whether to fire at them or not without any further intervention by human operators. In the supervised mode, firing actions are contingent on the judgment and the "go" command of a human operator.[21] Thus, the Super aEgis II satisfies the DoD requirement if it operates in the unsupervised

---

[17] See http://www.army-technology.com/projects/irondomeairdefencemi/ (accessed June 5, 2017)

[18] HRW, *Losing Humanity*, 12-13.

[19] See http://www.army-technology.com/projects/brimstone/ (accessed June 5, 2017).

[20] See http://www.dodaam.com/eng/sub2/menu2_1_4.php# (accessed June 5, 2017).

[21]   Simon   Parkin,   "Killer   Robots: The   soldiers   that   never   sleep",   *BBC*,   July   16,   2015 http://www.bbc.com/future/story/20150715-killer-robots-the-soldiers-that-never-sleep (accessed June 5, 2017).

mode, and it does not otherwise. In the latter case, it is best viewed as a combination of a decision-support system with a remote-controlled firing device. It is worth noting that the supervised Super aEgis II preserves most of the critical functions in the DoD requirement, insofar as it performs without any human intervention the tasks of target identification and tracking in its intended operational environment. Accordingly, this robotic sentinel affords a straightforward illustration of the fact that a simple on/off operational mode switch can make all the difference between an autonomous and a non-autonomous weapon system in the light of the DoD requirement on AWS as a binary discrimination criterion.[22]

By accepting the DoD requirements as a binary discrimination criterion, one is *ipso facto* committed to recognize the existence and actual deployment of several types of AWS, contrary to what was asserted by various representatives of State parties at CCW informal meetings of experts in Geneva between 2014 and 2016.[23] Moreover, it is worth noting that the distinctions between stationary/mobile or, more problematically, between defensive/offensive weapons systems are irrelevant to the problem of deciding whether a weapons system should be counted as autonomous or not on the basis of the DoD criterion. Indeed, Super aEgis II is a stationary system, and the Brimstone missile is not; the former is a defensive system and the latter is not. Fine-grained distinctions in terms of the types and levels perceptual and cognitive capabilities appear to be similarly immaterial with respect to this discrimination problem, insofar as the pattern matching capabilities that are required as a precondition for attacking a short-range rocket are quite different from the perceptual and cognitive capabilities that are required of a robotic sentry to recognize a human target or of an air-to-ground missile to identify armored enemy vehicles chiefly on the basis of their visual models.

Using the DoD requirements as an operational criterion enabling one to distinguish between autonomous and non-autonomous weapons systems is in some respects unsatisfactory, at least insofar as these requirements force one to place in the same AWS class both defensive systems like the Iron Dome, which exclusively reacts to and targets incoming short range rockets, and any future offensive system which will be supposedly capable to operate autonomously in the fog of war, and to discriminate active foes from civilians and *hors de combat* enemies in cluttered urban warfare scenarios. It was noted above that this inconvenience cannot be overcome by endorsing the

---

[22] Tamburrini, *On Banning AWS*, 126.

[23] In addition to the stance taken by the UK, mentioned above (note 10), this view was expressed, on the basis of equally strict requirements of autonomy, by Argentina (2014 CCW Meeting of Experts, Closing Statement), Denmark (2015 CCW Meeting of Experts, General Exchange), France (2016 CCW Meeting of Experts, Working Paper on "Characterization of a Lethal AWS"), Germany (2015 CCW Meeting of Experts, Final Statement), Italy (2016 CCW Meeting of Experts, Statement at the Panel "Towards a Working Definition of Lethal AWS"), Japan (2016 CCW Meeting of Experts, Working Paper), Sweden (2014 CCW Meeting of Experts, Opening Statement), and Turkey (2016 CCW Meeting of Experts, General Exchange).

prohibitively restrictive UK condition on autonomy. At the same time, in order to introduce a new definition of AWS striking a good balance between liberal and restrictive definitions, one has to address the formidable problem of identifying a non-arbitrary cutting point in between the perceptual, cognitive, and action capabilities of systems like the Iron Dome on the one hand, and those characterizing the envisaged AWS that will be capable of matching infantryman skills in cluttered urban warfare scenarios.

At a closer look, this question resembles the long-standing (and possibly unending) philosophical discussion as to what intelligence and intelligent behavior are in connection with the aims and achievements of Artificial Intelligence (AI), Cognitive Robotics and cognate disciplines. Yet, one cannot reasonably expect to make the establishment of an international regime for AWS conditional upon the solution of such theoretical conundrum. On a precautionary basis, therefore, the more liberal requirement put forth by the DoD - however imperfect - should be preferred as a basis for distinguishing between AWS and other weapons systems. In this way, it is possible to avoid deferring the discussion on AWS regulation to some undetermined technological future, if and when the highly restrictive requirements envisaged by the UK conditions will be met.

2.1 From "autonomy" to "(lack of) meaningful human control"

The uncertainties surrounding the meaning of "autonomy" in the expression "autonomous weapons systems" should not be overemphasized for our present purposes. In the current debate on AWS, the quest for a shared notion of autonomy has been gradually losing ground in favor of a (brand-new) definitional dispute concerning the expression "meaningful human control". Apparently, NGOs, diplomats and scholars are getting less and less interested in attaining precise operational definitions enabling one to decide, for each existing and developmental weapons system, whether it is autonomous or not. To be investigated, rather, is what kind or level of human control must on ethical and legal grounds be exercised on weapons systems.

The idea whereby all weapons (thus including AWS) should be subject to a meaningful human control was ushered in by Article 36, a UK-based NGO, in a 2013 Paper commenting upon the aforementioned UK Ministry of Defense's Joint Doctrine.[24] This formula, which was refined in subsequent policy briefs, turned out to be particularly successful and soon influenced the overall Campaign strategy. Significantly enough, in a 2014 report, Human Rights Watch has come to

---

[24] Article 36, *Killer Robots: UK Government Policy on Fully Autonomous Weapons*, April, 2013, http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf (accessed June 5, 2017).

define AWS as those which "would identify and fire on targets without meaningful human intervention".[25] Likewise, the 2015 Open Letter concludes with a call for a "ban on offensive autonomous weapons beyond meaningful human control".[26] The (lack of) meaningful human control is thereby embodied and plays a crucial role in the very definition of autonomous weapons, with the consequence that the campaign for banning AWS and that for introducing a MHC requirement end up pursuing partially overlapping goals. And clearly, the more restrictive is the interpretation one gives to the MHC requirement, the more inclusive is the class of lethal AWS that *ipso facto* ought to be prohibited. This point was vividly made by Kerstin Vignard, Deputy to the Director of the UN Institute for Disarmament Research (UNIDIR), when she observed at the 2016 CCW Informal Meeting of Experts that "[u]ltimately the autonomy question is really about what control/oversight do we expect humans to maintain over the tools of violence that we employ".[27]

As anticipated, the case for MHC, unlike the original pro-ban campaign, has rapidly attracted the interest of an increasing number of States, which have been using it as a viable starting point for the discussion on *Lethal* AWS at the CCW Informal Meetings of Experts.[28] As a consequence of this, it has been questioned whether the notion of MHC already gathered (or is going to gather in a near future) a sufficient consensus for giving birth to a new norm of customary international law.[29]

In fact, if framed in these terms, the search for a legal basis of a MHC requirement on autonomous weapons systems is bound to fail for (at least) three reasons. First, to date the notion of MHC has been expressly endorsed by little more than two dozens of States: this constitutes *at best* a manifestation of *opinio iuris* by a (non-negligible) group of States, but this is clearly insufficient to infer the existence - even at an inchoate stage - of a customary norm. Second, the evidentiary value

---

[25] Human Rights Watch, *Shaking the Foundations: The Human Rights Implications of Killer Robots*, May 12, 2014, 1 (hereinafter HRW, *Shaking the Foundations*).

[26] In a similar vein, see Maina Kiai and Christof Heyns, Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies, UN Doc. A/HRC/31/66, 4 February 2016, para. 67(f) ("Autonomous weapons systems that require no meaningful human control should be prohibited […]") and African Commission on Human and Peoples' Rights, *General Comment No. 3 on the African Charter on Human and Peoples' Rights: The Right to Life (Article 4)*, 57th Ordinary Session, 4-18 November 2015, para. 35 (hereinafter African Commission, *General Comment No. 3*) ("Any machine autonomy in the selection of human targets or the use of force should be subject to meaningful human control").

[27] Kerstin Vignard, *Statement of the UN Institute for Disarmament Research at the CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems*, April 12, 2016, http://www.unidir.org/files/medias/pdfs/unidir-s-statement-to-the-ccw-informal-meeting-of-experts-on-lethal-autonomous-weapon-systems-eng-0-648.pdf (accessed June 5, 2017).

[28] Dustin A. Lewis, Gabriella Blum, and Naz K. Modirzadeh, *War-Algorithm Accountability*, Harvard Law School Program on International Law and Armed Conflict, Research Briefing, August 2016, 62 (hereinafter Lewis, Blum and Modirzadeh, *War-Algorithm*). See also UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward*, 2014, http://www.unidir.ch/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf (accessed June 5, 2017).

[29] Peter Asaro, "Jus nascendi, robotic weapons and the Martens Clause" *in Robot Law*, Ed. by Ryan Calo et al. (Edward Elgar, 2016) 367.

of these declarations is somehow diminished by the fact that they mostly come from States which would not be "specially affected" by an international regime on AWS, as they would not be able in any case to develop this technology in a foreseeable future,[30] with the notable exception of the United Kingdom.[31] Third, the formation of a new norm of customary international law is fatally hindered by disagreements among States as to what the introduction of a MHC requirement would specifically entail.[32]

These difficulties should not lead one to relinquish the case for MHC, though. Indeed, they all derive from asking the wrong question at the outset: the point is not whether a rule imposing MHC on autonomous weapons has been emerging as a new customary norm; rather it should be investigated whether such a rule can be inferred from the *existing* international legal framework. Our contention is that it can be, building upon international humanitarian law, international criminal law, and international human rights law. Furthermore, we will argue that the very same ethical and legal considerations which substantiate the existence of an MHC requirement in international law provide a sound guidance as to the level of control that humans have to exercise on weapons systems.

In the next two sections, therefore, the main arguments originally framed in the AWS ban debate, and hence supporting the existence of an MHC requirement will be analyzed in turn, distinguishing - as mentioned above - those crafted in decidedly deontological terms from those having a more consequentialist orientation.

3. Deontological Perspectives on Banning AWS

A rich family of arguments for banning AWS are framed in deontological terms, with solid underpinnings in international law. Deontological arguments for banning AWS can be informatively classified into three groups on the basis of the claims that each one of them is meant to buttress. First, AWS would be unable to comply with various obligations flowing from both IHL and IHRL rules governing the use of lethal force. Second, to the extent that they take humans 'out-of-the-loop', AWS are likely to determine an accountability gap. And the latter is hardly

---

[30] Lewis, Blum and Modirzadeh, *War-Algorithm*, 58.

[31] In response to a joint letter from UK members of the Campaign to Stop Killer Robots, sent to the Foreign Secretary, a high-profile UK officer declared that "[i]t is the policy of the Government that the operation of our weapons will always be under human control as an absolute guarantee of human oversight, authority and accountability for weapons use." See Article 36, *UK government: Defining 'human control' essential at killer robots talks in 2017*, January 12, 2017, http://www.article36.org/autonomous-weapons/uk-govt-response-2017/ (accessed June 5, 2017).

[32] Rebecca Crootof, "A Meaning Floor for 'Meaningful Human Control'" *Temple International & Comparative Law Journal* 30(2016): 53, 55-56.

reconcilable with the agent-relative moral obligation of military commanders and operators to be accountable for their own actions, as well as with the related principle of individual criminal responsibility under ICL. Third, the deployment of AWS would run contrary to a central foundational value of both IHL and IHRL, namely human dignity and the patient-relative rights that come with it, which require that the taking of human life should be reserved to human decision-makers. Let us turn to examine the main reasons that have been adduced for each one of these claims.

3.1 AWS are Unable to Comply with IHL and IHRL Rules Governing the Use of Lethal Force

The main legal norms appealed to in deontological arguments for banning AWS are the IHL rules of distinction[33] and proportionality.[34] It is seriously doubted that AWS will be able, in the foreseeable future, to comply with these rules in a wide variety of warfare scenarios.[35] An AWS complying with IHL requirements is usually taken to be an autonomous weapon which is capable of respecting the principles of distinction and proportionality *at least as well as a competent and*

---

[33] Article 48, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) 1977, *entered into force* Dec. 7, 1978, 1125 U.N.T.S. 3 (hereinafter AP I) ("[…] the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives"). See also Articles 51, para. 2 ("The civilian population as such, as well as individual civilians, shall not be the object of attack. […]") and 52, para. 1 AP I ("Civilian objects shall not be the object of attack or of reprisals"). According to the codification of customary IHL rules edited by the ICRC, this rule has acquired the status of customary international law. See Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law*, Vol. I (Cambridge University Press, 2006), 3-8 (hereinafter ICRC's Study on Customary IHL).

[34] The rule of proportionality prohibits "[l]aunching an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated, is prohibited". See ICRC's Study on Customary IHL, 46. References to this rule may be found in Articles 51, para. 5(b) and 57 para. 2 (a)(iii) and (b) AP I.

[35] Noel Sharkey, "The Evitability of Autonomous Robot Warfare", *International Review of the Red Cross* (*IRRC*) 94 (2012): 787 (hereinafter Sharkey, *The Evitability of Autonomous Robot Warfare*); David Akerson, "The Illegality of Offensive Lethal Autonomy" *in International Humanitarian Law and the Changing Technology of War*, Ed. Dan Saxon (Brill, 2013) 65; *Heyns Report*, paras. 63-74; Chantal Grut, "The Challenge of Autonomous Lethal Robotics to International Humanitarian Law", *Journal of Conflict & Security Law* 18(2013): 5 (hereinafter Grut, *The Challenge of Autonomous Lethal Robotics*); Jeroen van den Boogaard, "Proportionality and Autonomous Weapons Systems", Journal of International Humanitarian Legal Studies 6(2015): 247 (hereinafter van Boogard, *Proportionality and AWS*); Kjølv Egeland, "Lethal Autonomous Weapon Systems under International Humanitarian Law", *Nordic Journal of International Law* 85(2016): 89 (hereinafter Egeland, *Lethal Autonomous Weapon Systems*). This issue has also been raised, although with different overtones, by many of the delegations intervened at the CCW Meetings of Experts. See the views expressed by Argentina (2014 CCW Meeting of Experts, Closing Statement), Austria (2015 CCW Meeting of Experts, Working Paper "The concept of 'meaningful human control'"), Cuba (2015 CCW Meeting of Experts, Working Paper, 2015), Ecuador (2016 CCW Meeting of Experts, General Exchange), the Holy See (2015 CCW Meeting of Experts, Statement at the Panel on Overarching Issues), Norway (2014 CCW Meeting of Experts, Opening Statement); Pakistan (2015 CCW Meeting of Experts, General Exchange), South Africa (2014 CCW Meeting of Experts, Opening Statement), Sweden (2014 CCW Meeting of Experts, Statement at the Panel on "Legal Aspects – International Humanitarian Law"), Switzerland (2016 CCW Meeting of Experts, Statement at the Panel on "Challenges to International Humanitarian Law"), and United Kingdom (2015 CCW Meeting of Experts, Statement at the Panel on "Challenges to International Humanitarian Law").

conscientious human soldier. The possibility of an AWS fulfilling IHL in this sense presupposes the solution of many profound research problems in advanced robotics and AI.[36]

On the one hand, while present-day UAVs "cannot distinguish a sleeping dog from a bush, even at high noon,"[37] the rule of distinction raises the problem of endowing AWS with the capability of perceptually distinguishing civilians and *hors de combat* people in erratic and surprise seeking warfare environments, in addition to the capability of establishing whether civilians have lost protection from attacks because of their participation into hostilities.[38] Thus, in particular, AWS must be capable of recognizing behaviors that conventionally or unconventionally carry hostile or surrender messages and fighting incapacitation information, on the basis of a variety of perceptual and cognitive capabilities, including viewpoint-independent recognition of bodily postures and gestures in the variable perceptual conditions of unstructured warfare scenarios, understanding of emotional expressions, and real-time reasoning about deceptive intentions and actions.

On the other hand, the proportionality rule requires the decision-maker to strike a delicate balance between military gains expectedly deriving from some given course of action and harms to civilians (in terms of life loss and injuries or damages to civilian objects) ensuing from it. In fact, it is argued, a proper appraisal of the terms of this balance, as well as of their interactions can hardly be carried out by present and foreseeable AI systems,[39] especially insofar as it requires advanced social interaction competence and high sensitivity to the emotional and cognitive context, in addition to the military commander's evaluation of human factors which include the shared viewpoints and feelings arising in soldiers on account of extended periods of communal daily life, training, and field experience.

This point can be further elucidated with an example. Let us consider a simplified military scenario where the decision-maker faces the following two options: Option A, which will yield substantial military advantage, is likely to provoke a significant amount of civilian casualties; and Option B, which will cause less damage to the enemy, is unlikely to harm civilians. A military commander who has decided that the proportionality principle is respected in option A might still refrain from

---

[36] Noel Sharkey, "Saying 'no!' to lethal autonomous targeting", *Journal of Military Ethics* 9(2010): 369, 378; Ronald Arkin, "Lethal autonomous systems and the plight of the non-combatant", *AISB Quarterly* 137 (2013): 1, 4 (hereinafter Arkin, *Lethal autonomous systems*).

[37] Lora G. Weiss, "Autonomous Robots in the Fog of War", *IEEE Spectrum*, July 27, 2011, http://spectrum.ieee.org/robotics/military-robots/autonomous-robots-in-the-fog-of-war (accessed June 5, 2017).

[38] Philip Alston, "Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law", *Journal of Law, Information and Science* 21(2011): 35, 54 (hereinafter Alston, *Lethal Robotic Technologies*); Sharkey, *The Evitability of Autonomous Robot Warfare,* 788-789; Grut, *The Challenge of Autonomous Lethal Robotics*, 11-12; Robin Geiss, *The International-Law Dimension of Autonomous Weapons Systems*, Friedrich Ebert Stiftung, October 2015, 14 (hereinafter Geiss, *The International-Law Dimension of AWS*).

[39] Sharkey, *The Evitability of Autonomous Robot Warfare*, 789-790; *Heyns Report*, paras. 70-73; Grut, *The Challenge of Autonomous Lethal Robotics,* 12-14; Egeland, *Lethal Autonomous Weapon Systems*, 103-105.

following this course of action on account of the nefarious impact which, knowing well her or his soldiers and sharing with them daily warfare hardships, she or he deems that the killing of civilians will have on the troops' morale. This negative effect will in fact, according to her or his best judgment, jeopardize in the long run the possibility of defeating the enemy in the given operational theatre. It is presently difficult to imagine technologically foreseeable machines which are capable of gauging the psychological state of the troops involved in a conflict and the related knock-on strategic effects. Accordingly, a military commander delegating in similar scenarios the task of appraising the implications of the proportionality rule and the ensuing operational decisions to an AWS would *ipso facto* violate an agent-relative obligation attached to her or his special role of commander in a warfare theatre.

Additional concerns have been raised in relation to the AWS' ability to comply with the rule of precaution in attacks, which supplements the aforementioned rules of distinction and proportionality by obligating warring parties to take all feasible steps to avoid, or at least to minimize, the harm caused to innocent civilians.[40] The focus, here, is basically on the concept of "feasibility" which, similarly to those of "civilians taking part in the hostilities" and "disproportionate collateral damage" referred above, requires complex contextual evaluations which are unlikely to be adequately performed in the foreseeable future by a non-human decision-maker.[41]

Similar arguments have been put forward with reference to IHRL, having particular regard to the right to life and to bodily integrity. Notably, it has been submitted that, should AWS be deployed outside the context of an armed conflict, as a means of law enforcement (e.g. to quell a riot or to control a border[42]), the thorny problems emphasized in relation to IHL would assume an even greater magnitude. Since the notions of "legitimate object of attack" and "collateral damages" are not contemplated under IHRL, lawful killings are considered there as an exception, subject to very strict conditions. Indeed, the requirements of necessity and proportionality, despite similarities in wording, are construed far more stringently than in IHL, which makes the hypothetical tasks of a law enforcement AWS even more difficult to fulfill.[43]

---

[40] Art. 57 AP I. On the customary nature of this principle, see ICRC's Study on Customary IHL, 51-67.

[41] Nathalie Weizmann and Milena Costas Trascasas, *Autonomous Weapon Systems under International Law*, Geneva Academy of International Law and Human Rights, Briefing No. 8 (2014), 16; Alston, *Lethal Robotic Technologies*, 54.

[42] Like the sentinel robotic guns mentioned in section 2. On this issue, see Maya Brehem, *Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law*, Geneva Academy of International Law and Human Rights, Briefing No. 9 (2014) (hereinafter Brehem, *Defending the Boundary*).

[43] Titus Hattan, "Lethal Autonomous Robots: Are They Legal under International Human Rights and Humanitarian Law?", *Nebraska Law Review* 93(2015): 1035, 1053-1054 (hereinafter Hattan, *Lethal Autonomous Robots*); Cristof Heyns, "Human Rights and the use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement", *Human Rights Quarterly* 38(2016): 350, 362-366 (hereinafter Heyns, *Human Rights and the use of AWS*). See also African Commission, *General Comment No. 3*, paras. 31 and 35.

3.2 Recourse to AWS determines an accountability gap

Even the more convinced proponents of the ethical acceptability and legality of AWS are compelled to admit that, no matter how accurate, these systems are error-prone in ways which are "completely alien to humans".[44] As a consequence, it is quite possible for an AWS to commit acts amounting – at least materially – to war crimes or crimes against humanity.[45] But then who will be personally responsible for these conducts? This question lies at the core of the second pro-ban argument, whereby *the delegation of the use of force to non-human decision-makers would create an accountability gap*.[46] This proposition stands in need of some further clarification. To begin with, it goes without saying that, because of the lack of moral agency, AWS could not be held responsible as direct perpetrators.[47] Responsibility for their actions, therefore, should be traced back to some persons in the decision-making chain. And that is where the real problems begin.

At the outset, one should take note that the list of potentially responsible individuals is quite long, as it includes 'the software programmer, [the producer of the AWS], the military commander in charge of the operation, the military personnel that sent the AWS into action or those overseeing its operation, the individual(s) who conducted the weapons review, or political leaders'.[48] Such a long list, far from facilitating the task of identifying the responsible individuals, is likely to raise the familiar 'many hands' problem. This problem commonly occurs in software-related accidents,[49] where a group of people can be held collectively responsible for a determined outcome, whereas none of them can often be individually blamed for it.[50]

To illustrate the many hands problem in the specific case of a war crime charge, consider the scenario of an autonomous UAV model which is programmed to distinguish between military and civilian buildings. Evidence collected in the UAV testing phase suggests that the system fails to identify as civilian objects buildings presenting certain perceptual features in 5 % of cases. In order

---

[44] Paul Scharre, *Autonomous Weapons and Operational Risk*, Ethical Autonomy Project, February 2016, 17.

[45] Rebecca Crootof, "War Torts: Accountability for Autonomous Weapons", *University of Pennsylvania Law Review* 164(2016): 1347, 1375-1377 (hereinafter Crootof, *War Torts*).

[46] See, generally, Human Rights Watch, *Mind the Gap: The Lack of Accountability for Killer Robots*, April 9, 2015 (hereinafter, HRW, *Mind the Gap*).

[47] Robert Sparrow, "Killer Robots", *Journal of Applied Philosophy* 24(2007): 62, 71 (hereinafter Sparrow, *Killer Robots*); Sharkey, *The Evitability of Autonomous Robot Warfare*, 790; Egeland, *Lethal Autonomous Weapon Systems*, 91.

[48] Markus Wagner, "Autonomous Weapon System", *in Max Planck Encyclopedia of Public International Law*, Ed. Rüdiger Wolfrum (Oxford University Press, 2016), margin number 21 (hereinafter Wagner, *AWS*).

[49] Helen Nissenbaum, "Accountability in a computerized society", *Science and Engineering Ethics* 2(1996): 25.

[50] On the many hands problem, see generally Dennis F. Thompson, "Moral Responsibility of Public Officials: The Problem of Many Hands", *The American Political Science Review* 74(1980): 905. This notion has been recently applied to international legal responsibility by André Nollkaemper, "The Problem of Many Hands in International Law" *in The Political Economy of International Law. A European Perspective*, Ed. Alberta Fabbricotti (Edward Elgar, 2016) 278.

to encourage governmental purchase, however, the manufacturer falsifies the test results, by lowering the failure rate to 0,5 %. When delivering the purchased system to the battlefield unit, the Secretary of Defense omits to mention the failure rate, solely recommending the greatest caution in the use of the aircraft against targets in the proximity of civilian objects. Basing her or his judgment on a liberal interpretation of this directive, a military commander deploys the aircraft over a sparsely populated area. As a consequence of this chain of events, a hospital is inadvertently bombed and several civilian victims are counted.

Each character mentioned in this story contributed in distinctive ways to the accident occurrence: clearly, their acts are significant causal antecedents of the bombing event; more significantly, the conduct of each involved individual is reprehensible for some reason: the programmer should not have released the software in the face of a significant failure rate; the manufacturer should not have falsified the test results; the Secretary of Defense should have been clearer as to the risks associated with use of the AWS; the military commander should have been more cautious. And yet, none of them, taken individually, can be reasonably deemed to be responsible as such (both directly and indirectly) for the bombing of the hospital. Accordingly, it would be unfair to charge anyone of them with a war crime.

Accountability gaps may also occur beyond the (somewhat convoluted) "many hands" scenario. Clearly, there may be uncontroversial cases such as that of a machine which was deliberately pre-programmed to carry out international crimes or that of a commander who deployed an AWS in a context different from the one it was designed for, and where it was likely to commit war crimes.[51] In the majority of conceivable cases, however, the complexities of AWS technologies and their behavioral unpredictability in partially structured or unstructured warfare scenarios are likely to afford a powerful defense against criminal prosecution. Indeed, since AWS might be capable of taking courses of action whose reason "may be opaque even to the system's designers",[52] there will be cases where it is impossible to ascertain the existence of the mental element (intent, knowledge or recklessness[53]), which is required under ICL to ascribe criminal responsibilities. As a

---

[51] Crootof, *War Torts*, 1377.

[52] *Ibidem*, 1373.

[53] It should be noted that it is fairly controversial whether "recklessness" constitutes a culpable state of mind under customary international law. One should note, in this respect, that "recklessness" was not included as a culpable *mens rea* under Article 30 of the Statute of the International Criminal Court, *entered into force* July 1, 2002, 2187 U.N.T.S. 90 (see, critically, Antonio Cassese, "The Statute of the International Criminal Court: Some Preliminary Reflections", *European Journal of International Law* 10(1999): 144, 154).

consequence, no one could be held criminally liable, notwithstanding the conduct at stake clearly amounts to a war crime or a crime against humanity.[54]

The doctrine of "command responsibility" is not particularly helpful either in order to fill in this accountability gap. An essential element of this doctrine is the commander's effective control over the subordinate's behavior. This element can be hardly applied here. On the one hand, the very notion of "autonomy" implies the absence of (or, at least, a limited role for) human supervision. On the other hand, AWS' faster-than-human reaction times would make commander's control to a large extent a purely speculative option.[55] Military commanders, in other words, would be stripped of the agent-relative moral duties that are traditionally associated with their rank and roles, namely, the duty to assess the risk of violating IHL requirements and to set out their orders accordingly.[56]

In sum, the deployment of AWS in warfare (and, possibly, law enforcement) scenarios gives rise to the serious possibility that international crimes are perpetrated, but no individual is legally responsible for those outcomes. This is in stark contrast with a "cornerstone" of contemporary international law,[57] the principle of individual criminal responsibility.[58] The importance of this principle cannot be overestimated. Contrary to what some authors seem to suggest,[59] the crucial, two-fold function of *deterring* the commission of international crimes and adequately *retributing* the offender for the harm done is peculiar to ICL and cannot be performed in the same way by the law of State responsibility. As famously noted by the Nuremberg Tribunal, in fact, international

---

[54] Sparrow, *Killer Robots*, 69-70; Grut, *The Challenge of Autonomous Lethal Robotics*, 14-17; HRW, *Mind the Gap*, 19-20; Geiss, *The International-Law Dimension of AWS*, 20-21; Egeland, *Lethal Autonomous Weapon Systems*, 110-112; Crootof, *War Torts*, 1375-1378. See also, with specific reference to Active Protective Systems, Michael Kurt Riepl, "War crimes without criminal accountability? The case of Active Protection Systems", *Humanitarian Law & Policy*, June 1, 2016, http://blogs.icrc.org/law-and-policy/2016/06/01/war-crimes-without-criminal-accountability-case-active-protection-systems (accessed June 5, 2017).

[55] See, also for further references, Crootof, *War Torts*, 1380.

[56] Tamburrini, *On Banning AWS*, 131-132. For a more general account of the expected negative impact of AWS on "military virtue", see Shannon Vallor, "The Future of Military Virtue: Autonomous Systems and the Moral Deskilling of the Military", *in 2013 5th International Conference on Cyber Conflict*, Ed. by Karlin Podins, Jan Stinissen and Markus Maybaum (NATO CCD COE Publications, 2013) 471.

[57] Christian Tomuschat, "The Legacy of Nuremberg", *Journal of International Criminal* Justice 4(2006): 830, 840.

[58] A considerable number of States have raised this issue at the CCW Meetings of Experts. See the views set forth by Argentina (2015 CCW Meeting of Experts, General Exchange), Austria (2014 CCW Meeting of Experts, Opening Statement), Chile (2016 CCW Meeting of Experts, Statement at the Panel on "Challenges to International Humanitarian Law"), Cuba (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), Ecuador (2016 CCW Meeting of Experts, General Exchange), France (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), the Holy See (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), Mexico (2016 CCW Meeting of Experts, General Exchange), Norway (2016 CCW Meeting of Experts, Statement at the Panel on "Challenges to International Humanitarian Law"), Pakistan (2014 CCW Meeting of Experts, Opening Statement), Poland (2015 CCW Meeting of Experts, Statement at the Panel on "Characteristics of the Lethal AWS"), Spain (2015 CCW Meeting of Experts, General Exchange), Sri Lanka (2015 CCW Meeting of Experts, General Exchange), and Switzerland (2015 CCW Meeting of Experts, Statement at the Panel on "Challenges to International Humanitarian Law").

[59] Crootof, *War Torts*, 1386-1402.

crimes "are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced".[60]

3.3 AWS' lethal decision-making runs contrary to human dignity

The upshot of the third pro-ban argument is that the delegation of lethal decision-making to a machine would be prohibited under the principle of human dignity[61] (or, somehow relatedly, under the "principles of humanity" enshrined in the so-called Martens Clause).[62] This argument is more far-reaching than the previous ones, as it leverages on the principle of human dignity, which is both foundational and open-textured.[63] This argument has been described as being "at the core of the concerns raised about fully autonomous weapons".[64] In fact, the very idea of a machine endowed with the power to take life-or-death decisions is intuitively "repugnant".[65] While the basic claim underlying this argument is straightforward ("there is a violation of dignity when a machine kills a human being"), it is informative to distinguish analytically between two of its variants, one of which is centered on agent-relative duties, and the other one on patient-relative rights.

---

[60] International Military Tribunal Nuremberg, *The Trial of German Major War Criminals*, Judgment, 1 October 1946. Proceedings of the International Military Tribunal sitting at Nuremberg, Germany, Part 22 (22nd August, 1946 to 1st October, 1946), at 447. The Norwegian delegate at the 2016 CCW Meeting of Experts put the problem in very similar terms: "Another intrinsic challenge with fully autonomous weapons would be ensuring individual […] responsibility for unlawful acts in times of armed conflict. This is a cornerstone of modern international law. Without accountability, deterring and preventing international crimes becomes all that much harder" (Statement at the Panel on "Challenges to International Humanitarian Law").

[61] At the CCW Meetings of Experts, this argument was put forth with particular vigor by Germany (2014 CCW Meeting of Experts, Opening Statement) and the Holy See (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"). See also the stance taken by Chile (2015 CCW Meeting of Experts, General Exchange), Costa Rica (2016 CCW Meeting of Experts, Statement at the Panel on "Human Rights and Ethical Issues"), Croatia (2014 CCW Meeting of Experts, Closing Statement), Cuba (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), Ecuador (2014 CCW Meeting of Experts, Opening Statement), France (2015 CCW Meeting of Experts, General Exchange), Morocco (2016 CCW Meeting of Experts, General Exchange), Mexico (2015 CCW Meeting of Experts, Statement at the Panel on "Transparency and the Way Ahead"), Pakistan (2015 CCW Meeting of Experts, Opening Statement), Sierra Leone (2016 CCW Meeting of Experts, General Exchange), Sweden (2015 CCW Meeting of Experts, General Exchange), and Switzerland (2014 CCW Meeting of Experts, Opening Statement).

[62] Article 1, para. 2 AP I "[…] civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, *from the principles of humanity* and from the dictates of public conscience" (emphasis added). The issue of compatibility of robotic killings with the Martens Clause was raised at the CCW Meetings of Experts by the following States: Austria (2015 CCW Meeting of Experts, Working Paper on "The concept of 'meaningful human control'"), Brazil (2014 CCW Meeting of Experts, Opening Statement), Ecuador (2016 CCW Meeting of Experts, General Exchange), and Mexico (2014 CCW Meeting of Experts, Opening Statement).

[63] Heyns, *Human Rights and the use of AWS*, 367. See also, in a critical perspective, Dieter Birnbacher, "Are autonomous weapon systems a threat to human dignity?", *in Autonomous Weapons Systems: Law, Ethics, Policy*, Ed. Nehal Bhuta et al. (Cambridge University Press, 2016), 105.

[64] UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values*, 2015, 7, http://www.unidir.org/en/publications/the-weaponization-of-increasingly-autonomous-technologies-considering-ethics-and-social-values (Accessed June 5, 2017).

[65] *Ibidem*, 8.

The argument based on human dignity was first put forward by Peter Asaro in a leading article published in 2012 on the International Review of the Red Cross.[66] In its original formulation, this argument emphasizes agent-relative duties flowing from the principle of human dignity. The starting point of this argument is that the action of suppressing a human life is legally or morally justifiable only if it is non-arbitrary. However, for an act of taking human lives to be non-arbitrary, the agent should ground his/her action 'on a considered and informed decision'.[67] That is where the principle of human dignity comes in. The latter, in fact, dictates that the act of killing be based on *human* judgment, for only human decision-making guarantees the full appreciation "of the value of individual life [and] the significance of its loss".[68] Therefore, to the extent that lethal decision-making is carried out by a machine without any involvement of human judgment, AWS are inherently unable to comply with this dictate. Accordingly, AWS taking human life are neither legitimate nor morally justifiable.[69]

The variant of the argument based on human dignity which is centered on patient-relative rights was clearly formulated by the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns.[70] His argument is grounded in the Kantian conception of human dignity, which entails – ultimately – that people have the inherent right to be treated as "complete and unique human beings", especially when their lives are at stake. Accordingly, human dignity would be blatantly denied if people were subject to robotic lethal decision-making, because this would place them "in a position where an appeal to the humanity of the person on the other side is not possible". Indeed, the decision to kill or not would be taken "on the basis of hypotheticals", which would be set in advance, in the AWS programming phase, or would be developed by the machine itself as rules of behavior extrapolated from its past experience, provided that the machine had been suitably

---

[66] Peter Asaro, "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making", *International Review of the Red Cross* 94(2012): 687.

[67] *Ibidem*, 689.

[68] HRW, *Shaking the Foundations*, 3.

[69] It was aptly noted that the prohibition flowing from the principle of human dignity might be independently endorsed from a consequentialist standpoint in normative ethics on account of its expected consequences as a behavioral rule. Indeed, by requiring the involvement of human judgment in life-or-death decision-making may selectively induce a human agent, whose motives comprise compassion and empathy, to refrain from using lethal force, notwithstanding she or he is legally entitled to do so (Geiss, *The International-Law Dimension of AWS*, 18). If compassion and empathy do no act as a restraint, "greater killing and suffering in war than would otherwise be necessary" could ensue (Adam Saxton, "(Un)Dignified Killer Robots? The Problem with the Human Dignity Argument", *LawfareBlog*, Mar. 20, 2016, https://www.lawfareblog.com/undignified-killer-robots-problem-human-dignity-argument (accessed June 5, 2017). And, in turn, if AWS "always apply deadly force within what is legally permissible the possibility exists that people will die who otherwise would have been spared" (Geiss, *The International-Law Dimension of AWS*, 18). At the CCW Meetings of Experts, this point was made, effectively enough, by the delegate of the Holy See, who observed that:
> "Studies of soldiers' experiences support that human beings are innately averse to taking life, and this aversion can show itself in moments of compassion and humanity amidst the horrors of war. Programming an 'ethical governor' or 'artificial intelligence' to enable autonomous weapon systems to technically comply with the law of war in the areas of distinction and proportionality, even if possible, is not sufficient." (2014 CCW Meeting of Experts, Opening Statement).

[70] Heyns, *Human Rights and the use of AWS*, 370-371.

endowed with some learning algorithm for this purpose. The ensuing death-or-life decision could hardly be overridden when the AWS is about to actually release force, with the consequence that the human target would be somehow "written off" without the (even slightest) hope of changing his/her fate. In this way, hope, which constitutes a consubstantial part of human nature, would become "one of the casualties of AWS".[71]

3.4 The scope of deontological arguments against AWS

Deontological arguments suffer from some inherent limitations in scope, which must be duly taken into account in order to assess their capability to support the case for a comprehensive ban on AWS.

First, while no educated guess can be sensibly advanced as to whether and when AWS will meet IHL, one cannot *ipso facto* exclude the technological possibility of IHL-compliant AWS.[72] Accordingly, deontological arguments drawing on current and foreseeable violations of IHL by AWS do offer a strong support for a moratorium on the development and use of lethal AWS *in those warfare scenarios where significant problems of distinction and proportionality arise*.[73] But they are not sufficient to support an unconditional ban on AWS. For all we know, lack of IHL compliance is a contingent situation which may no longer hold at some future time, and the same may be said with regard to IHRL.

Second, the arguments concerning the AWS-related accountability gap do not take into account scenarios where war crimes and crimes against humanity cannot simply be committed, such as aerial and submarine duels between AWS taking place with no humans in their range of action or AWS' attacks directed solely against military communication satellites and other uninhabited military infrastructures. Similarly, the Asaro/Heyns argument addresses problems arising from the *lethal* use of AWS only. Therefore, these arguments offer support for the introduction of a prohibition to employ AWS in those warfare scenarios, which are admittedly crucial for ethical and

---

[71] In a similar vein, see Brehem, *Defending the Boundary*, 65 ("treating the programming of algorithms as an adequate implementation of legal obligations ignores the objectifying and dehumanizing potential of autonomous targeting on a procedural level. In this reading, the calculated blindness to individual circumstances involved in the use of an AWS is an affront to human dignity").

[72] At the CCW Meetings of Experts, this point was highlighted by several EU States: Czech Republic (2014 CCW Meeting of Experts, Statement at the Panel on "Operational and Military Aspects"), Finland (2015 CCW Meeting of Experts, Statement at the Panel on "Transparency and the Way Ahead"), France (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), Greece (2015 CCW Meeting of Experts, Statement at the Panel on "Possible Challenges to IHL"), Italy (2016 CCW Meeting of Experts, Statement at the Panel on "Towards a Working Definition of Lethal AWS"), and Spain (2014 CCW Meeting of Experts, Opening Statement).

[73] Arkin, *Lethal autonomous systems,* 7. See also Michael N. Schmitt and Jeffrey S. Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict", *Harvard National Security Journal* 4:(2013) 231; Hattan, *Lethal Autonomous Robots*, 1048-1051.

legal reflections on war, where human lives are at stake. However, they are silent about, say, AWS attacking uninhabited infrastructures or about warfare scenarios involving AWS only.[74]

To sum up. A distinction was introduced between deontological arguments which depend on the failure of present and foreseeable AWS to comply with IHL and deontological arguments which do not depend on this contingency, but rather hinge on the need to avoid accountability gaps and to respect the right to human dignity. Deontological arguments of both kinds concern what human agents of certain special sorts that are involved in warfare scenarios morally owe to other human beings that are involved in the same scenarios as potential patients of aggressive actions, in addition to the related question whether these moral and legal obligations would be necessarily or likely to be violated using AWS against human beings. Therefore, these arguments do not bear directly on the problem of prohibiting the use of AWS against uninhabited military infrastructures and in actions against other AWS. Accordingly, there are some limitations of deontological arguments for banning AWS, as they are solely concerned with AWS lethal uses.

Some of these deontological concerns might be selectively defused by assigning a role to human operators "in" or "on" the decision loop through the introduction of an (aptly crafted[75]) MHC requirement on lethal weapons systems. The fact remains, however, that the spread of AWS, regardless of the level of human control to which their lethal uses are subject, is likely to yield destabilizing effects at a global and regional scale. These moral and legal problems, which lie outside the scope of the deontological arguments examined so far, are tackled by consequentialist arguments for an AWS ban, insofar as these latter arguments concern non-lethal uses of these weapons too.

4. Consequentialist perspectives on banning AWS

Consequentialist approaches in normative ethics assess moral obligations and permissions exclusively on the basis of an evaluation of the (actual or expected) consequences of actions. Accordingly, consequentialist and deontological appraisals of AWS deployment may conflict with each other. Indeed, depending on the criteria that one adopts to evaluate action consequences, some AWS attacks on humans may be permissible in view of actual or expected consequences that are deemed to be good; and AWS attacks on uninhabited military objectives – which may be morally permissible from the deontological viewpoint of agent-relative obligations and patient-relative

---

[74] See *infra* section 6.
[75] Schmitt, *Autonomous Weapon Systems*, 11 ("[n]ot every battlespace contains civilians"). See also van Boogard, *Proportionality and AWS*, 262.

rights – might be forbidden instead from a consequentialist viewpoint.[76] Thus, the range of moral judgments that one may advance from a consequentialist perspective differs in significant ways from the range of moral judgments one may offer on the basis of the deontological arguments examined in the previous section. In the next sub-sections, we will focus on consequences of AWS deployment with respect to various conceptions of what are good and bad consequences in the realm of armed conflicts and wars. Again, as in the case of deontological arguments, normative ethics and international law approaches intersect. In this case, however, the crucial intersections concern more directly the legal regime governing the maintenance of peace and security rather than IHL and IHRL.

4.1 Narrow vs. wide consequentialist arguments

To begin with, let us note that any consequentialist argument for or against a ban on AWS is presently bound to focus on expected, rather than actual, consequences of their deployment since only relatively unsophisticated AWS (see section 2 above) from a much wider spectrum of possibilities are in actual use.[77] Thus, the consequentialist debate about an AWS ban is more appropriately framed as a *rule-consequentialist* debate concerning which one of two incompatible rules (prohibiting or permitting AWS) must be adopted by the international community of state actors in view of their *expected* outcomes. Accordingly, the debate has essentially developed in terms of different appraisals of the expected consequences of AWS future deployment. In this connection, an informative distinction can be made between wide and narrow appraisals of the expected consequences of AWS deployment.[78]

Narrow approaches to the appraisal of expected consequences concentrate on expected AWS battlefield performances and some of their outcomes. Wide approaches take into account, in addition to local battlefield implications, expected geopolitical consequences of AWS deployment, which range from regional destabilization and conflicts to global implications on arms races and

---

[76] In our opinion, the *2016 Summer Study on Autonomy* by the US Science Defence Board does not provide a sufficiently comprehensive picture of arguments against AWS, insofar as it seems to belittle the role of consequentialist arguments against AWS that are not restricted to their lethal uses. Indeed, it is claimed there that "Skepticism about the employment of autonomy in military operations is almost wholly focused on the use of autonomous weapons systems with the potential for lethality." (https://www.hsdl.org/?view&did=794641, p. 20. Accessed June 5, 2017).

[77] It is interesting to mention, in this connection, the opinion of the Office of the US Air Force Chief Scientist: "[i]t is possible to develop systems having high levels of autonomy, but it is the lack of suitable V&V [validation and verification] methods that prevents all but relatively low levels of autonomy from being certified for use". See Office of the US Air Force Chief Scientist, *Technology Horizons: A Vision for Air Force Science and Technology 2010–30*, vol. 1, AF/ST-TR-10-01, May 15, 2010, ix.

[78] Tamburrini, *On banning AWS*, 137-141.

conflict deterrence.[79] Thus, wide consequentialist approaches differ from narrow consequentialist approaches in that they evaluate AWS deployment in the light of its expected impact over and above battlefield operations. More specifically, wide consequentialist approaches take into account lower disincentives to start war, a new arms race, regional and global destabilization threats – going up to and including nuclear deterrence destabilization.

Those who defend the rule that AWS should be permitted on consequentialist grounds have usually assumed a narrow perspective on expected consequences. Roughly speaking, from this perspective AWS development and deployment should be permitted insofar as these new conventional arms are expected to bring about reduced casualties in one's own and the opponents' camp, as well as among non-belligerents who happen to be present on the battlefield. This expectation is grounded in the belief that AWS will be capable of performing more accurate targeting than human soldiers, and will be programmed to adopt more conservative decisions to fire insofar as these machines can be made free from human self-preservation concerns.[80] The force of these narrow consequentialist arguments for the future deployment of AWS depends on another crucial assumption. This is the *ceteris paribus* assumption that the deployment of AWS will not have a significant impact outside battlefield scenarios. However, the weakness of this *ceteris paribus* assumption has been convincingly and repeatedly brought out.[81] Indeed, one may reasonably expect that the spreading of AWS will bring about comprehensive and long-term consequences for international security along with local and short-term military advantages on the battlefield. A document produced by ICRAC (International Committee for Robot Arms Control)[82] summarizes various threats to international security raised by AWS concentrating in particular on the proliferation of these weapons with oppressive regimes[83] and terrorists,[84] their mass proliferation among state actors giving rise to a

---

[79] A wide consequentialist approach is pursued by, e.g., Jürgen Altmann, "Arms control for armed uninhabited vehicles: an ethical issue", *Ethics and Information Technology* 15(2013): 137; and Noel Sharkey, "Why robots should not be delegated with the decision to kill", *Connection Science*, 29(2017): 177, 182-183.

[80] See, generally, Ronald Arkin, *Governing Lethal Behavior in Autonomous Robots* (CRC Press, 2009).

[81] This weakness is aptly emphasized, e.g., in the 2015 Open Letter and by Stuart Russell, "Take a Stand on AI Weapons", *Nature* 521(2015): 415.

[82] ICRAC, *LAWS: Ten Problems for Global Security.* Memorandum for delegates at the 2015 CCW Meeting of Experts, 13-17 April 2015, http://icrac.net/wp-content/uploads/2015/04/LAWS-10-Problems-for-Global-Security.pdf (accessed June 5, 2017).

[83] HRW, *Losing Humanity*, 38 ("Fully autonomous weapons would conversely be perfect tools of repression for autocrats seeking to strengthen or retain power. Even the most hardened troops can eventually turn on their leader if ordered to fire on their own people. A leader who resorted to fully autonomous weapons would be free of the fear that armed forces would rebel. Robots would not identify with their victims and would have to follow orders no matter how inhumane they were"); Marco Sassòli, "Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified", *International Law Studies* 90(2014): 308, 317 (hereinafter Sassòli, *Autonomous Weapons*) ("if autonomous weapons fall into the hands of a ruthless leader able to program them, they will never abandon his or her cause as soldiers might because the programmed robots will be unable to come to the conclusion that it is unjust, even when there are repeated and widespread violations of IHL").

[84] Serious concerns about the possible proliferation of AWS among terrorist groups were expressed at the CCW Meetings of Experts. See the observations made by Algeria (2016 CCW Meeting of Experts, General Exchange),

new arms race,[85] less disincentives to start wars,[86] on account of the reduced numbers of soldiers that will be involved, and correspondingly lowered thresholds for armed conflicts, unpredictability of interaction with friendly or enemy AWS, their cyber vulnerability possibly leading to unintended conflicts, acceleration in the pace of battle, in addition to continuous global battlefields brought about by AWS left behind to patrol post-conflict zones over long time periods.

In addition to these concerns, one should carefully note that future AWS, more than many other conventional arms, have the potential to deliver destructive attacks on nuclear objectives. Large swarms of AWS flying at supersonic and hypersonic speeds might be capable of delivering a powerful first strike against the opponent's nuclear arsenals, to the extent that they may thwart the opponent's capability of responding with nuclear retaliation. In this scenario, nuclear deterrence based on mutually assured destruction would no longer count as a motivation to withhold aggression and first strike strategies would be prized instead.[87]

What is then a wide consequentialist appraisal of the overall expected benefits and costs flowing from AWS deployment? Arguably, by permitting the AWS deployment, one might expect the good consequence of reduced casualties among belligerents and non-belligerents in some local battlefield scenarios. By taking this course of action, however, one would significantly raise at the same time the danger of starting a new arms race leading to regional and global destabilization risks, up to and including the weakening of traditional nuclear deterrence factors based on mutually assured destruction. As the latter negative consequences outweigh the sum of the expected benefits flowing from AWS deployment, the collective rule of behavior that is expected to produce the preferable set of consequences in a global geopolitical context is that of prohibiting – rather than permitting – the production and deployment of AWS.

4.2 Consequentialist pro-ban arguments in a legal perspective

In the framework of international law debates, consequentialist approaches have been equally pursued by those backing or else opposing a ban on AWS. Significantly enough, the two sides of

---

Austria (2014 CCW Meeting of Experts, Opening Statement), Cuba (2015 CCW Meeting of Experts, General Exchange), Ecuador (2016 CCW Meeting of Experts, General Exchange), the Holy See (2016 CCW Meeting of Experts, Working Paper "Elements Supporting the Prohibition of Lethal Autonomous Weapons Systems"), Pakistan (2014 CCW Meeting of Experts, Opening Statement), and Sierra Leone (2016 CCW Meeting of Experts, General Exchange).

[85] At the CCW Meetings of Experts, this risk has been underscored by Algeria (2016 CCW Meeting of Experts, General Exchange), Austria (2015 CCW Meeting of Experts, General Exchange), Cuba (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), and the Holy See (2016 CCW Meeting of Experts, Working Paper "Elements Supporting the Prohibition of Lethal Autonomous Weapons Systems").

[86] See *infra* section 4.2.

[87] See Tamburrini, *On banning AWS*, 139-140.

the debate are positioned coherently with the distinction, set out above, between narrow and wide consequentialist arguments. In fact, the consequentialist approach is commonly wielded by the anti-ban front, which has argued that AWS' deployment will ultimately result in 'higher-than-human' performances with respect to adherence to IHL, because robots can become more accurate than human soldiers in targeting military objectives and, unlike human soldiers, are utterly unconstrained by the need for self-preservation and immune from human passions (such as anger, fear and vengefulness).[88]

As noted above, however, this narrow appraisal only captures a fraction of the overall picture, since it is confined to the battlefield-related effects and screens off (by the implicitly assumed *ceteris paribus* clause) more pervasive effects that are likely to flow from AWS deployment. Indeed, supporters of a ban reach opposite conclusions on the basis of a broader consideration of the consequences that one may expect from an increased use of AWS. This enlarged perspective brings into play a distinct legal regime, since one moves from the law regulating the conduct of hostilities (IHL, or *jus in bello*) to the law pertaining to the maintenance of international peace and security (Art. 1, para. 1 UN Charter). The latter includes, but is not limited to, the rules governing the use of force, or *jus ad bellum* (viz. the prohibition on the use of force under Art. 2 para. 4 UN Charter; the right to self-defence under Art. 51 UN Charter; and the collective security system governed by Chapter VII of the UN Charter).[89]

A preliminary observation on the shifting legal framework is in order here. While the potential impact of AWS on international peace and security is often described as a matter of concern for international law, this proposition has been rarely discussed in depth, so that is not entirely clear what is the actual legal issue at stake. This led one author to radically rule out the relevance of *jus ad bellum* in this field, in the light of the fact that the determination as to '[w]hether a breach of a rule of *ius ad bellum* has occurred […] is independent from the type of weapon that has been used'.[90]

The latter view is not without foundation. Whether a certain use of force is contrary to the *jus ad bellum* ultimately depends on the circumstances in which force is unleashed (Who? Against whom? Why?) and not on the sorts of weapons that are employed. If, for instance, State A deploys a swarm

---

[88] Schmitt, *Autonomous Weapon Systems*, 23; Kenneth Anderson and Matthew C. Waxman, "Law and Ethics for Autonomous Weapon Systems. Why a Ban Won't Work and How the Laws of War Can", *Columbia Public Law Research Paper* 11(2013): 8, 14; Sassòli, *Autonomous Weapons*, 310.

[89] Peter Asaro, "How just could a robot war be?" in *Current Issues in Computing And Philosophy*, Ed. by Adam Briggle, Katinka Waelbers and Philip A.E. Brey, (Ios Press 2008) 50, 56-58; Alston, *Lethal Robotic Technologies*, 44; HRW, *Losing Humanity*, 39-41; Grut, *The Challenge of Autonomous Lethal Robotics*, 22-23; Geiss, *The International-Law Dimension of AWS*, 12-13.

[90] Wagner, *AWS*, marginal number 11.

of AWS against State B, the legality of its conduct will be gauged, under *jus ad bellum*, on the basis of the following elements: whether State A acted in self-defense or whether the use of force was authorized by the UN Security Council. Conversely, it will be completely immaterial whether the attack was carried out through AWS or other alternative means.

Yet, and again, this is not the whole story. The law governing the maintenance of international peace and security cannot be reduced to a static, binary decision rule. This legal regime is not only about determining whether a specific armed activity is lawful or not under the prohibition on the use of force. Rather, it is about ensuring – in the words of the 1984 Declaration on the Right of Peoples to Peace – that 'the policies of States be directed towards the elimination of the threat of war'.[91] This claim entails that a more comprehensive (and dynamic) appraisal must be carried out, which may well include an evaluation of policies allowing the use of AWS, especially in connection with the question whether these policies are conducive to more peace and security in international relations or, on the contrary, represent a factor of instability at global and regional levels. Should the latter be the case, any such policy would not only be undesirable as a matter of normative ethics, but also as a matter of international law, as it would run contrary to the maintenance of international peace and security, namely, according to a commonly shared view, it would run counter to the 'purpose of all purposes' of the UN Charter.[92]

In the previous sub-section, a variety of nefarious wide-scale consequences were listed, which are likely to ensue from permissive policy towards AWS. Each one of these consequences, taken individually, is arguably sufficient to support the contention that an AWS permissive policy should be outlawed as detrimental to the achievement of the UN goal of a world order of peace and security. Here, we limit ourselves to underlining that a policy allowing the use of AWS would end up encouraging a more liberal approach to the use of force by States. In their turn, such liberal approaches may bring about a higher likelihood of violations of the prohibition on the use of force under Article 2 para. 4 of the UN Charter.

The present argument is based on a straightforward assumption. Even if one does not fully embrace the (admittedly controversial) democratic peace theory, one must acknowledge that, in democratic countries, public opinion and legislative assemblies play an important role in deterring governments from deploying their armed forces in aggressive military campaigns. In this respect, a crucial factor lies in the risk of casualties among national military personnel. Indeed, popular outrage generally stemming from the return of "flag-draped coffins" represents a significant incentive for

---

[91] U.N.G.A. Resolution 39/11 (1984), para. 3.
[92] Rüdiger Wolfrum, "Ch. I Purposes and Principles, Article 1" in *The Charter of the United Nations: A Commentary*, 3rd ed., Ed. by Bruno Simma et al. (Oxford University Press, 2012) 108, 109.

representatives sitting in parliaments to exert a meaningful control over the use of war powers by the executive. As a collateral (and not necessarily intended) effect, this democratic dynamics may prevent States from breaching the prohibition on the use of force. A notable case in point occurred in 2013, when the US and UK governments renounced to wage (an arguably unlawful) war against Assad in Syria apparently in view of the disapproval expressed by domestic public opinion and parliamentary representatives.[93]

A policy allowing for the use of AWS would inevitably affect this (obliquely) virtuous circle. If human troops are replaced, say, by robots, the potential cost of the conflict in terms of human losses significantly decreases (when it does not equate to zero) and, with it, sensitivity to the issue in the general public. Accordingly, legislative assemblies would be less motivated to control the governmental exercise of war powers, thereby encouraging further executive unilateralism in this field.[94] As a final result, democratically unchecked military operations will be more and more likely to occur, leading to more frequent breaches of the prohibition on the use of force. As the Austrian delegation openly put it at the 2014 CCW Meeting of Experts: "[p]utting soldiers' lives at stake makes States think twice whether to engage in armed conflict. Autonomous weapons remove such restraint from the use of military force".[95]

Significantly enough, a strikingly similar conclusion has been reached by Rebecca Crootof, a resolute critic of an AWS ban. In a recent analysis devoted to the influence of AWS on the distribution of war powers in the US legal system, Crootof describes how (and why) the development of this technology would lead to a further concentration of the war power in the Executive's hands.[96] When turning to consider the international legal implications of this process, she notes that, as a result of this concentration of power, US Presidents "will be more willing to

---

[93] Hélène Dieck, *The Influence of Public Opinion on Post-Cold War U.S. Military Interventions* (Palgrave MacMillan, 2015) 185; Andrew Grice, "Syria crisis: The British public has its say as two-thirds oppose strikes", *Independent*, Sep. 3, 2013, http://www.independent.co.uk/news/uk/politics/syria-crisis-the-british-public-has-its-say-as-two-thirds-oppose-strikes-8795319.html (accessed June 5, 2017).

[94] For discussion of this issue, especially in connection with the use of drones in the US 2011 Libya intervention, see Edoardo Datteri and Guglielmo Tamburrini, "Robotic Weapons and Democratic Decision-Making" in *Robotik und Gesetzgebung*, Ed. by Eric Hilgendorf and Jan-Philipp Günther (Nomos, 2013) 211.

[95] 2014 CCW Meeting of Experts, Opening Statement. It is of the utmost significance that the same concern has been recently voiced by the Chinese delegation in a position paper submitted at the CCW Fifth Review Conference (see http://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China's+Posit ion+Paper.pdf; accessed June 5, 2017). See also the views expressed by Costa Rica (2016 CCW Meeting of Experts, General Exchange), Chile (2015 CCW Meeting of Experts, General Exchange), Cuba (2015 CCW Meeting of Experts, General Exchange), Ecuador (2016 CCW Meeting of Experts, General Exchange), France (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), the Holy See (2015 CCW Meeting of Experts, Statement at the Panel on "Overarching Issues"), Pakistan (2014 CCW Meeting of Experts, Opening Statement), and Sri Lanka (2015 CCW Meeting of Experts, General Exchange).

[96] Rebecca Crootof, "War, Responsibility, and Killer Robots", *North Carolina Journal of International Law and Commercial Regulation* 40(2015): 909.

engage in humanitarian interventions".[97] At this juncture, the author acknowledges that unilateral humanitarian interventions are still prohibited under international law, but suggests that a more frequent resort to military force by the US could lead to the consolidation of a new exception to the prohibition on the use of force under Article 2, para. 4 of the UN Charter.[98] Whether this result could ever be achieved, however, it is highly doubtful, if one only considers that substantial portions of the world community have treated US-led unilateral humanitarian interventions with the utmost suspicion. What is less controversial, for the time being, is that more frequent resort to military force will ultimately mean more frequent violations of the prohibition on the use of force. Accordingly, Crootof's statement quoted above is aptly turned on its head and rephrased as follows: because of the concentration of power stemming from the use of AWS, US Presidents "will be more willing to engage in *violations of the prohibition on the use of force under Art. 2 para. 4 UN Charter*". This amended conclusion affords a quite strong motivation for banning AWS!

C. The scope of consequentialist arguments against AWS

If taken in isolation, consequentialist arguments – similarly to deontological ones – suffer from some inherent limitations, which must be duly taken into account in order to assess their force in support of a comprehensive ban on AWS.

First, AWS proliferation with oppressive regimes and terrorist groups could be effectively countered by the introduction of an *ad hoc* control legal regime (either in the form of a treaty or of a Security Council resolution), aimed at strengthening international cooperation to prevent AWS from ending up in the "wrong" hands.[99]

Second, not every use of AWS by States is likely to yield destabilizing effects. Just consider, for instance, AWS performing merely defensive (e.g. the Israeli Iron Dome or certain Active Protective Systems) or sentinel functions (such as the gun turret installed along the demilitarized zone between South and North Korea). In this perspective, one might argue, a comprehensive ban would appear as an excessive legal response. Indeed, States and NGOs might alternatively negotiate a prohibition on the production and use of AWS for aggressive and potentially destabilizing purposes, leaving aside, e.g., those carrying out defensive and sentinel functions, which represent a substantive portion of AWS that are currently in use.

---

[97] *Ibidem*, 929.
[98] *Ibidem*, 930-931.
[99] Sassòli, *Autonomous Weapons*, 317.

In summary, by sufficiently enlarging the temporal and spatial horizon of what must be included in the list of expected consequences, it appears that a policy allowing, in general terms, for the use of AWS is likely to produce negative destabilizing effects that largely outweigh positive ones. However, this argument, if taken in isolation, provides limited support only for banning AWS, since the expected destabilizing effects deriving from an increasing spread and use of AWS might be successfully offset through the adoption of various restrictive measures falling short of a sweeping ban.

5. Deontological and consequentialist reasons for banning AWS: a confluence model

Let us recap the main points of the analysis developed in the previous sections.

- Deontological arguments for the conclusion that AWS are unable to comply with the rules governing the use of lethal force in both IHL and IHRL depend on technological limitations affecting current and foreseeable AWS. As a consequence, they effectively support limited actions only, towards the establishment of a moratorium or the development of special regulation on AWS permissible uses.

- In contrast with this, the remaining deontological arguments (i.e. the "accountability gap" and "human dignity" arguments) are independent of technologically contingent situations, and make rather appeal to ethically and legally entrenched obligations of special sorts of agents or to some patient-relative human rights. Accordingly, these arguments are found to effectively support only a prohibition of *lethal AWS*, but afford weaker and only indirect arguments for AWS not targeting humans.[100] Moreover, moral and legal concerns about lethal AWS might be adequately addressed by selectively introducing suitable MHC requirements and without outlawing AWS *as such*.

- Wide consequentialist arguments raise strong objections against AWS, and also against those AWS that are not directed towards human targets, *as long as there are expected destabilizing effects at various scales*. However, if no destabilizing effects are expected, and good consequences are in the offing without being counterbalanced by bad consequences, consequentialist arguments may lead one to support the introduction of AWS for targeting humans in view of their greater targeting accuracy and more conservative firing decisions.

---

[100] In a "slippery slope" perspective, for instance, it may argued that, if the use of AWS not targeting humans is allowed, the possibility remains that a commander could feel compelled, out of military necessity, to employ such AWS in a scenario involving the targeting of human beings.

Accordingly, there is a tension between deontological and consequentialist approaches to the issue of AWS deployment and use, which leads to conflicting conclusions about lethal uses of AWS having no evident destabilizing effects. There is, however, elbow room for reconciling these different approaches[101] thereby enlarging the scope of AWS ban requests.

To begin with, let us notice that deontological arguments in normative ethics concern some agent-relative obligations and patient-relative rights in warfare scenarios involving the use of *lethal* force by means of AWS. The conclusions of deontological arguments against AWS notably flow from inviolable foundational values (human dignity) and requirements of moral responsibility and accountability that are deeply entrenched into IHL, IHRL and ICL. Under these bodies of law, rules concerning the use of lethal force are categorical and cannot be derogated from in the pursuit of a greater good (as a consequentialist analysis would suggest), except for special excusing conditions that are explicitly envisaged by the rules themselves.

In fact, while international law is certainly permeable to the influence of consequentialist thinking, some of its norms posit inviolable precepts, whose respect is not amenable to trade-offs against utilitarian considerations. Reference is made to *jus cogens* norms, the (Kantian[102]) deontological *noyau dur*, sitting at the top of the hierarchy of the sources of international legal order. To the extent that the basic tenets of IHL, IHRL and ICL rank as *jus cogens* (indeed, a relatively unproblematic assumption)[103], therefore, deontological arguments based upon them take precedence over consequentialist ones *as a matter of positive international law*,[104] and thus independently from any philosophical stand one may wish to take on the "deontology vs. consequentialism" ethical tensions.

Accordingly, the joint ethical and legal reinforcement of categorical obligations motivates the following Prioritization of Deontology rule (PD-rule): *the conclusions of deontological arguments against AWS cannot be overridden in any circumstance in which they are applicable*. However, it was noticed that there are circumstances in which deontological arguments do not provide such guidance, i.e. when the obligations of certain sorts of agents or the rights of certain sorts of potential patients are not at stake. In these circumstances, deontological approaches – both ethical and legal ones – are inapplicable. How is this deontologically unregimented space of action possibilities to be

---

[101] Michael S. Moore, "Patrolling the borders of consequentialist justifications: the scope of agent-relative restrictions", *Law and Philosophy* 27(2008): 35, 38-44.

[102] Sten Verhoeven, "Norms of Jus Cogens in International Law: A Positivist and Constitutionalist Approach" (Phd Diss., Katholieke Universiteit Leuven, 2011), 314-341.

[103] It is not possible to engage here in a careful demonstration of this assertion. It suffices to recall, to this end, the first two reports on *jus cogens* by ILC's Special Rapporteur Dire Tladi (Un Doc. A/CN.4/693, Mar. 8, 2016; UN Doc. A/CN.4/706 Mar. 16, 2017) which contain several references pointing at that direction.

[104] Indeed, even an author who is avowedly in favour of a consequentialist approach to international justice, is compelled to concede that "international law treats *jus cogens* norms as trumps". See Steven R. Ratner, *The Thin Justice of International Law: A Moral Reckoning of the Law of Nations* (Oxford University Press 2015) 81.

dealt with? Here, the agent-neutral (and patient-neutral) consequentialist framework can provide the required guidance, insofar as wide consequentialist reasons agree with and are enshrined into the principles (set forth by the UN Charter and related legal instruments) concerning peace and security. In the deontologically unregimented space of action, wide consequentialist reasons are reinforced by a variety of legal instruments concerning peace and security. Thus, deontological and consequentialist frameworks can be amalgamated by assigning them to different domains, thereby avoiding intertheoretical conflicts in normative ethics and legal conflicts between different bodies of law. In other words, one is led to added to the PD-rule the following default rule (DEF-rule): apply consequentialist arguments against AWS whenever deontological arguments are inapplicable.

In virtue of its merging of deontological and consequentialist frameworks, the confluence model based on the prioritization of deontological reasons against AWS (PD model from now on) bolsters the case for an extensive AWS prohibition over and above *lethal* AWS, as it includes consequence-oriented reasons to curb some potential uses of AWS that have no direct lethal effects on human beings. A recently discussed scenario of this kind concerns autonomous Unmanned Underwater Vehicles (UUV), which may be used for trailing ballistic missile submarines. In a British Pugwash report by Sebastian Brixey-Williams it is claimed that "within a decade …adaptable long-endurance or rapidly-deployable unmanned underwater vehicles (UUV) and unmanned surface vehicles (USV), look likely to undermine the stealth of existing submarines".[105] Autonomy appears to be a crucial feature for the task that these UUV are supposed to perform, insofar as remote control signals do not travel well in salted underwater environments. And clearly, there is no direct lethal or sublethal effect on human beings produced by autonomous UUV which compromise the stealth of ballistic missile submarines. However, implementing this scenario may have a significant impact on peace and stability. Indeed, the difficulty of detecting submarines fitted with acoustic quieting systems makes them survivable candidates in the case of a first nuclear strike. By the same token, these vessels are assigned a key role in the MAD nuclear deterrence strategy. Therefore, undermining their stealth by means of autonomous UUV may undercut a major condition for the mutual acceptance of the MAD strategy. In these circumstances, the prioritized deontological rule does not apply, but a request for prohibiting these uses of autonomous UUV can be supported by appealing to the default consequentialist rule of the PD model.

6. The confluence model and interpretations of meaningful human control

---

[105] Sebastian Brixey-Williams, "Will the Atlantic become transparent?", *British Pugwash*, November 2016, http://www.basicint.org/sites/default/files/Pugwash_TransparentOceans_update_nov2016_v1(1).pdf (accessed June 5, 2017)

As anticipated, the proposed PD model bears some important implications for the problem of assessing what it is to exercise a genuinely meaningful human control on the all-important subclass of AWS formed by *lethal* weapons systems, an issue which appears to be crucial in the drafting of a future ban treaty on Lethal AWS.

The starting point, which is shared also by the opponents of a ban, is that "humans will decide under what parameters and frameworks it will operate, and when and where it will be deployed".[106] Controversy arises, however, when we move to consider the human-machine interaction in relation to individual attacks. To illustrate the implications of the PD-confluence model in this respect, let us first recall the distinction that Noel Sharkey[107] introduced between five different levels of human control on weapons systems:

1. A human engages with and selects targets, and initiates any attack;

2. A program suggests alternative targets and a human chooses which to attack;

3. A program selects targets and a human must approve before attack;

4. A program selects targets and a human has restricted (but still manageable)[108] time to veto;

5. A program selects targets and initiates attack without human involvement.

We contend that the issue of MHC does not lend itself to a one-size-fits-for-all solution that can be achieved, e.g., by endorsing one and only one of (1)-(5).[109] Nevertheless, what is the 'appropriate levels of human control'[110] in each case must be determined on a principled ethical and legal basis. One principled basis is afforded by appeal to the prioritized deontological rule of the PD model. In this respect, it could be useful to recall some classes of existing weapons systems mentioned in section 2 above, which satisfy the DoD requirements for autonomy:

---

[106] Jeffrey S. Thurnher, "Means and Methods of the Future: Autonomous Systems" *in Targeting: The Challenges of Modern Warfare*, Ed. by Paul A.L. Ducheine, Michael N. Schmitt, and Frans P.B. Osinga (T.M.C. Asser Press, 2016) 177, 181.

[107] Noel Sharkey, "Staying in the loop: human supervisory control of weapons" *in Autonomous Weapons Systems: Law, Ethics, Policy*, Ed. by Nehal Bhuta et al. (Cambridge University Press, 2016), 23, 34-37.

[108] The reference to "manageability" constitutes a slight (and yet important) deviation from Sharkey's scheme, insofar as it is aimed at ruling out from level 4 those circumstances, incurred by human operators of Iron Dome and similar reactive systems, where the time frame for overriding the system's decision is so short as to make human vetoing powers utterly illusory.

[109] The need for a flexible approach to MHC was advocated by Sweden at the 2015 CCW Meeting of Experts (General Exchange).

[110] This expression was ushered in by the US delegation as an alternative to a rigid notion of MHC (2016 CCW Meeting of Experts, General Exchange). See also DoD Directive, 2 ("It is DoD policy that: a. Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force").

A. STATIONARY SENTRIES. Autonomous sentinel gun towers, such as the South Korean Super aEgis II and the Turkish border sentinel, could be deployed at level 5 human control. But this possibility runs counter to the prioritized deontological rule of the PD model. The remaining human control levels, up to and including level 4, are not ruled out by the PD model. However, we suggest that level 3 is the appropriate MHC level, since the system intended operational conditions introduce no conflicting requirements for the protection of friendly forces: deployment environments and temporal constraints are fully compatible with those deliberation processes that one needs for ensuring human approval of sentry action.

B. FIRE-AND-FORGET LOITERING MUNITIONS. Any unleashing of the UK Brimstone (and similar fire-and-forget munitions[111]) must be subject to level 4 human control. This requirement can be further specified as follows. Human operators must provide in advance of each single activation an informed judgment about the kinds of targets that the system can attack; the operational area and the time-frame of action must be carefully recognized as ones where, according to the operator's pondered assessment, the fire-and-forget munitions system will not violate distinction and proportionality requirements. Moreover, human operators must be capable of vetoing the attack in the temporal window after activation and before strike, if evidence accrues that the environment is going to change so rapidly as to raise IHL (or IHRL) issues.

C. REACTIVE ANTI-MATERIEL DEFENSIVE SYSTEMS. In the case of Israeli Iron Dome (and similar defensive systems like the German Nächstbereichschutzsystem MANTIS), time-frame operational constraints do not allow for real-time judgments by human operators (even though they are *in abstracto* empowered to override the decision taken by the software). However, the system is neither directed against humans, nor has expected destabilizing effects. Thus, there are no constraints on autonomy levels flowing from either rule of the PD-model. Level 5 human control is allowed, provided that the system operates in areas preset by military officers on the basis of IHL-informed judgments[112]. If the AWS hits human targets, the military officer who decided to deploy the system, as well as

---

[111] For instance, the anti-radiation drones Harpy and Harpy 2 (or Harop) developed by the Israel Aerospace Industries which raise special concerns in view of their loitering time extending over a few hours, during which situational changes may occur that would require mission cancellation in the light of IHL and IHRL.

[112] Despite their eminently defensive character, a more nuanced solution is suggested in relation to Active Protective Systems (APS) for vehicles. This is because, unlike stationary systems, APS are intended for use in cluttered scenarios too, where issues of distinction and proportionality may well arise in their attack mode. Should it be the case, the MHC requirement would dictate to switch off the "fully autonomous mode" (level 5) in favor of - at least - level 4 human control. A shift from "autonomous" to "manual" mode is enabled by some APS that are already in use, like the South-African/Swedish LEDS-150 (http://saab.com/land/force-protection/self-protection/leds; accessed June 5, 2017).

programmers and manufacturers, can be held responsible for their culpable actions or omissions.

Agent-relative obligations and patient-relative rights, which constitute the prioritized backbone of the PD model, set constraints on the appropriate level of autonomy in cases A and B – by requiring the presence of humans formulating appropriate competent judgements and taking informed decisions for applying or vetoing each use of *lethal* force. In case C, no prescription issues from the PD model insofar as these AWS are neither directed against humans nor raise serious destabilizing concerns.

7. Concluding remarks

In the light of the above concrete applications of the PD-model, let us identify and emphasize what are the major legal outputs of our confluence model.

In the first place, the deontological prong of the model provides both the ethico-legal basis for a ban on Lethal AWS without meaningful human control and the way forward for a principled approach to the determination of appropriate levels of human control, which is not straitjacketed by fixed and decontextualized interpretations of what MHC amounts to. Accordingly, the prioritized deontological component of this model offers some food for thought for the attendees of the Group of Governmental Experts on Lethal AWS, which will meet from 13 to 17 November 2017 as well as, ideally, for the negotiators of a treaty banning Lethal AWS without meaningful human control.

In the second place, the PD model, taken as a whole, bolsters the case for an extensive AWS ban, which would be in two main respects wider than a ban proposal solely based on deontological reasons. On the one hand, it would reach beyond potentially lethal uses of these weapons systems, so as to cover also (future) AWS with expected destabilizing effects on international peace and security, even when they have no direct lethal or sublethal effects on human beings. On the other hand, the consequentialist prong of the model would support a ban on destabilizing AWS (be them lethal or not), regardless of what level of MHC is exercised on them. Admittedly, given the resistances currently encountered by proposals to outlaw Lethal AWS without meaningful human control, an even wider ban could come across as a rather far-fetched goal. Nevertheless, the PD model affords a valuable argumentative tool for NGOs campaigning for a ban on AWS, which is neither limited to their lethal uses nor circumscribed by the notion of MHC.

In conclusion, the normative PD model, by merging in a coherent fashion different ethical and legal frameworks, offsets the largely complementary weaknesses and takes advantage of the respective

strengths of deontological and consequentialist reasons that have been offered in public debates and scholarly contributions on the prohibition of AWS deployment and use. As we have tried to show, the proposed model operates more effectively than deontological and consequentialist reasons taken in isolation from each other; but also more effectively than a superficial juxtaposition of deontological and consequentialist reasons which disregards inter-theoretical tensions in normative ethics and consistency issues between different bodies of law.

---

[*] Professor of International Law, Dipartimento di Giurisprudenza, Università di Cagliari, Italy (daniele.amoroso@unica.it)

[†] Professor of Philosophy of Science and Technology, DIETI – Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Università di Napoli Federico II, Italy (tamburrini@unina.it).