GENEVA ACADEMY
Académie de droit international humanitaire et de droits humains
Academy of International Humanitarian Law and Human Rights

ACADEMY BRIEFING N°9

DEFENDING THE BOUNDARY

CONSTRAINTS AND REQUIREMENTS ON THE USE OF AUTONOMOUS WEAPON SYSTEMS UNDER INTERNATIONAL HUMANITARIAN AND HUMAN RIGHTS LAW

MAYA BREHM - MAY 2017

THE ACADEMY A JOINT CENTER OF
THE GRADUATE INSTITUTE GENEVA
INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DEVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES
UNIVERSITÉ DE GENÈVE
FACULTY OF LAW

DEFENDING THE BOUNDARY

CONSTRAINTS AND REQUIREMENTS ON THE USE OF AUTONOMOUS WEAPON SYSTEMS UNDER INTERNATIONAL HUMANITARIAN AND HUMAN RIGHTS LAW

ACKNOWLEDGEMENTS

This Academy Briefing was researched and written by Maya Brehm, Researcher at the Geneva Academy of International Humanitarian Law and Human Rights.

With thanks to Neil Davison, Merel Ekelhof, Sandra Krähenmann, Richard Moyes, Michael Siegrist and Ozlem Ulgen who provided helpful comments on an earlier draft, and to Munizha Ahmad-Cooke for her meticulous copy-editing.

The Geneva Academy would like to thank the Swiss Federal Department of Foreign Affairs (FDFA) for its support to the Geneva Academy's research on this issue.

DISCLAIMER

This Briefing is the work of the author. The views expressed in it do not necessarily reflect those of the project's supporters or of anyone who provided input to, or commented on, drafts. The designation of states or territories does not imply any judgement by the Geneva Academy or the FDFA regarding the legal status of such states or territories, or their authorities and institutions, or the delimitation of their boundaries, or the status of any states or territories that border them.

CONTENTS

KEY MESSAGES

- Autonomous weapon systems (AWS) tend to be portrayed as 'weapons of war', but international humanitarian law (IHL) would never be the sole, and in many instances, it would not be the primary legal frame of reference to assess the legality of their use. Consideration of international human rights law (IHRL) requirements and constraints on the use of AWS must be a part of the debate on AWS, including in the framework of the 1980 Convention on Certain Conventional Weapons (CCW).

- Where IHL permits the 'categorical' targeting of security measures, including the use of force, there is scope for the lawful use of an AWS. However, due to procedural requirements and the need to individuate the use of force, this scope is extremely limited under IHRL. IHRL requirements and constraints apply to the use of an AWS in an armed conflict in so far as they are not displaced by IHL.

- To safeguard human dignity and human rights, human agents must:
  - exercise the control necessary to determine, in a timely manner, what legal rules govern applications of force by means of an AWS, and adapt operations as required
  - remain involved in algorithmic targeting processes in a manner that enables them to explain the reasoning underlying algorithmic decisions in concrete circumstances
  - be continuously and actively (personally) engaged in every instance of force application outside of the conduct of hostilities
  - exercise active and constant (continuous or at least frequent, periodic) human control over every individual attack in the conduct of hostilities; they must appropriately bound every attack in spatio-temporal terms so as to enable them to recognize changing circumstances and adjust operations in a timely manner

1. INTRODUCTION

'[W]hen the lethal decision is purely automatic, the only human agent directly identifiable as the efficient cause of death would turn out to be the victim …' (G. Chamayou)[1]

'Law depends on violence and uses it as a counterpunch to the allegedly more lethal and destructive violence situated just beyond law's boundaries. But the violence on which law depends always threatens the values for which law stands.' (A. Sarat)[2]

Over recent years, there has been growing debate about the ethical, humanitarian, legal and security implications of autonomous weapon systems (AWS). The basic idea is that once activated, such weapon systems would detect, select and attack targets without further human intervention. According to leading researchers in the field of artificial intelligence (AI) and robotics, AI technology has 'reached a point where the deployment of such systems is — practically if not legally — feasible within years'.[3] AWS are said to have the potential to revolutionize warfare (and policing, although that argument is seldom made). Whilst success in the quest for AI may bring unprecedented benefits to humanity, it is also argued to pose an existential threat to humankind.[4]

A small number of states are actively engaged in research and development with the stated goal of increasing autonomy in weapon systems. Regarding the drivers for this trend, commentators cite a perceived need to react to threats more quickly, process growing data much more efficiently (speeding up the targeting–decision cycle), improve performance in communications-denied environments, increase persistence and endurance, and reduce the exposure of states' own security forces to physical harm.[5]

1 G. Chamayou, *Drone Theory*, Penguin Books, 2015, p 211.

2 A. Sarat, 'Situating Law Between the Realities of Violence and the Claims of Justice: An Introduction', in A. Sarat (ed), *Law, Violence, and the Possibility of Justice*, Princeton University Press, 2001, p 3.

3 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers', Future of Life Institute, 28 July 2015, http://futureoflife.org/open-letter-autonomous-weapons/.

4 S. Russell, D. Dewey and M. Tegmark, 'Research Priorities for Robust and Beneficial Artificial Intelligence', *AI Magazine* (2015) 105–114, http://futureoflife.org/data/documents/research_priorities.pdf.

5 'The changing character of war, fleeting nature of targets, and glut of big data requires the military to integrate machine learning into its targeting process to win wars' (C. Lewis, 'Capturing Flying Insects: A Machine Learning Approach to Targeting', *War on the Rocks*, 6 Sep 2016, http://warontherocks.com/2016/09/capturing-flying-insects-a-machine-learning-approach-to-targeting/). Others have cautioned that increasing automation exacerbates, rather than addresses challenges linked to speed and data load, and have pointed out that the protection of a state's own forces can be enhanced without autonomy in critical functions. See, e.g., United Nations Institute for Disarmament Research (UNIDIR), *Framing Discussions on the Weaponization of Increasingly Autonomous Technologies*, 2014, pp 5–6, http://www.unidir.ch/files/publications/pdfs/framing-discussions-on-the-weaponization-of-increasingly-autonomous-technologies-en-606.pdf. For a critical appraisal of the claim that AWS would help reduce cost, see R. R. Hoffmann, T. M. Cullen and J. K. Hawley, 'The Myths and Costs of Autonomous Weapon Systems', 72 *Bulletin of the Atomic Scientists* 4 (2016) 247–255, http://tandfonline.com/doi/abs/10.1080/00963402.2016.1194619?journalCode=rbul20.

The use of AWS can be expected to change the manner in which and the processes by which human beings exercise control over the use of force and its consequences. Out of concern over serious negative ethical, humanitarian, legal and security implications, policy makers and commentators have emphasized that human beings must retain 'meaningful', 'appropriate' or 'effective' control over weapons.[6] What that involves, concretely, remains to be clarified.

From a legal perspective, the requirement for meaningful human control over AWS would seem to entail that human agents involved in the use of an AWS have the opportunity and capacity to assess compliance with applicable legal norms and to take all legally required steps to respect and ensure respect for the law, including preventive and remedial measures. In what circumstances this is no longer the case was a point of contention among participants in informal expert meetings on 'lethal autonomous weapon systems' held in the framework of the 1980 Convention on Certain Conventional Weapons (CCW),[7] a UN treaty aiming to 'prohibit or restrict further the use of certain conventional weapons' in order to promote disarmament and the 'codification and progressive development of the rules of international law applicable in armed conflict'.[8]

At present, no rule of international law specifically prohibits or restricts the use of autonomy in weapon systems.[9] There is general agreement among CCW States Parties that '… any use of force, including through [AWS], must strictly comply with international law and, in times of armed conflict, with IHL'.[10] States Parties also agree that 1977 Additional Protocol I (AP I) to the Geneva Conventions imposes a legal obligation to determine whether the use of an AWS as a 'new weapon, means or method of warfare' would in some or all circumstances be prohibited under international law.[11]

Views diverge, however, on the circumstances in which it would be legal to use an AWS. The former UN Special Rapporteur on extrajudicial killings, Philip Alston, among others, noted the difficulty that military personnel face in present practice to distinguish between those who may be lawfully targeted and those who may not – 'decision-making [that] requires the exercise of judgement, sometimes in rapidly changing circumstances and in a context which is not readily susceptible of categorization'.[12] Alston's successor, Christof Heyns, elaborated on the challenges involved in translating context-dependent, value-based judgements implicit in the application of law into algorithms, and cautioned that taking human deliberation out of life-and-death decisions could be incompatible with human dignity and the principle of humanity.[13] Alston called on the international community to give urgent consideration to the 'ways in which proactive steps can be taken to ensure that [robotic] technologies are optimized in terms of their capacity to promote more effective compliance with international human rights and humanitarian law'.[14]

Participants in ongoing multilateral policy discussions are divided on whether 'proactive steps' should involve legally binding measures at the international level. UN Special Rapporteurs have argued that to the extent that AWS are not 'capable of complying with the requirements of IHL',[15] or, more broadly, that they 'require

6 Article 36, Structuring Debate on Autonomous Weapons Systems, Memorandum for Delegates to the Convention on Certain Conventional Weapons (CCW), 14–15 November 2013, http://www.article36.org/wp-content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf; M. C. Horowitz and P. Scharre, Meaningful Human Control *in Weapon Systems: A Primer*, Working Paper, Center for a New American Security (CNAS), March 2015, https://s3.amazonaws.com/files.cnas.org/documents/Ethical_Autonomy_Working_Paper_031315.pdf; International Committee of the Red Cross (ICRC), *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*, Report, Expert Meeting, Versoix, Switzerland, 15–16 March 2016, August 2016, p 7, https://shop.icrc.org/publications/autonomous-weapons-systems.html; Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) submitted by the Chairperson (Advanced Version), April 2016, §15, http://www.unog.ch/80256EDD006B8954/(httpAssets)/DDC13B243BA863E6C1257FDB00380A88/$file/ReportLAWS_2016_AdvancedVersion.pdf; 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; United States Department of Defense (US DoD), Directive no 3000.09, 21 November 2012, s 4(a), http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

7 The CCW held informal expert meetings on 'Lethal Autonomous Weapon Systems' in May 2014, chaired by France, and in April 2015 and April 2016, chaired by Germany. In December 2016, States Parties decided to formalize these discussions and established a Group of Governmental Experts 'related to emerging technologies in the area of lethal autonomous weapons systems (LAWS) in the context of the objectives and purposes of the Convention', to be chaired by India (Final Document (Advance Version), Fifth Review Conference of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, UN doc CCW/CONF.V/10, 23 December 2016.

8 Preamble, 1980 Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects.

9 US DoD, *Law of War Manual*, June 2015, s 6.5.9.2, p 329, http://www.dod.mil/dodgc/images/law_war_manual15.pdf.

10 'CCW Meeting of Experts: Possible Challenges to International Humanitarian Law Due to Increasing Degrees of Autonomy', Statement by Switzerland, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 13–17 April 2015, 15 April 2015. See also 'Towards a "Compliance-Based" Approach to LAWS', Informal Working Paper submitted by Switzerland, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 11–15 April 2016, 30 March 2016, §8, http://www.unog.ch/80256EDD006B8954/(httpAssets)/D2D66A9C427958D6C1257F8700415473/$file/2016_LAWS+MX_CountryPaper+Switzerland.pdf. In addition, other weapon treaties may apply. As pointed out in the US DoD *Law of War Manual*, supra fn 9, s 6.5.9.2, p 329, 'to the extent a weapon system with autonomous functions falls within the definition of a "mine"… it would be regulated as such'.

11 Art 36, 1977 Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (AP I). Arguably, a corresponding duty exists under customary IHL, binding all parties to an armed conflict (ICRC, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, 2006, p 4). For a discussion, see V. Boulanin, *Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapon Systems*, SIPRI Insights on Peace and Security no 2015/1, Stockholm International Peace Research Institute (SIPRI), November 2015, https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf.

12 Interim Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, UN doc A/65/321, 23 August 2010, §39.

13 E.g., Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, UN doc A/HRC/23/47, 9 April 2013, §§63–74 and 89–97; Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, UN doc A/71/372, 2 September 2016, §§75–83.

14 UN doc A/65/321, supra fn 12, §48.

15 UN doc A/HRC/23/47, supra fn 13, §63.

no meaningful human control',[16] they should be prohibited. A global civil society coalition is campaigning for a preventive ban on 'killer robots',[17] a call supported by some governments and scientists.[18] Others have rejected a ban as 'misguided',[19] arguing that more limited restrictions or non-legally binding measures (such as guidance on legal reviews, best practices or a manual on IHL interpretation) would be more appropriate,[20] or consider that the existing legal framework is sufficient to 'accommodate' issues raised by increasing autonomy in weapon systems.[21]

16 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, §67(f).

17 The Campaign to Stop Killer Robots, http://www.stopkillerrobots.org/.

18 See, e.g., 'Inputs by Pakistan', CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 13–17 April 2015, http://unog.ch/80256EDD006B8954/(httpAssets)/14636F3813F314DBC1257E21005A2A95/$file/LAWSPaper_Pakistan_CCW.pdf; 'Elements Supporting the Prohibition of Lethal Autonomous Weapons Systems', Working Paper submitted by the Holy See, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 11–15 April 2016, 7 April 2016, http://bit.ly/2eKacwv; International Committee for Robot Arms Control, 'The Scientists' Call to Ban Autonomous Lethal Robots', http://icrac.net/call/.

19 K. Anderson, D. Reisner and M. Waxman, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems', 90 *International Law Studies* (2014) 395; M. N. Schmitt and J. S. Thurnher, '"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict', 4 *Harvard National Security Journal* (2013) 233.

20 K. Anderson and M. Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, A National Security and Law Essay, Hoover Institution, Stanford University, 2013, p 23, http://media.hoover.org/sites/default/files/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf.

21 See, e.g., Statement by the United Kingdom of Great Britain and Northern Ireland, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 11–15 April 2016, http://unog.ch/80256EDD006B8954/(httpAssets)/37B0481990BC31DAC1257F940053D2AE/$file/2016_LAWS+MX_ChallengestoIHL_Statements_United+Kingdom.pdf, expressing confidence that IHL will be 'capable of dealing with an evolution in automation' as it has 'successfully accommodated' previous evolutions in military technology. For a compilation of states' positions as reflected in their statements at the 2015 and 2016 CCW meetings of experts, see Appendix II to D. A. Lewis, G. Blum and N. K. Modirzadeh, *War-Algorithm Accountability*, Research Briefing, Harvard Law School Program on International Law and Armed Conflict, August 2016, p 150, http://blogs.harvard.edu/pilac/files/2016/09/War-Algorithm-Accountability-Appendices-Only-Searchable-August-2016.pdf.

2. ABOUT THIS STUDY

With a view to supporting multilateral discussions on potential regulatory measures aimed at ensuring compliance with and promoting international legal norms that safeguard humanity, this study aims to deepen the understanding of the requirements and constraints that international legal standards for the protection of the human person place on the use of force by means of an AWS.

There is a rich and rapidly growing body of literature addressing many legal questions raised by AWS. The focus of scholarly inquiry has been on compliance with IHL rules on the conduct of hostilities, in particular, key rules on targeting, such as the rule on distinction. Comparably little attention has been given to the impact of AWS on human rights protection.[22] This is probably because many commentators and policy makers envision the use of AWS in the context of military combat, rather than policing, and because discussions within the CCW are limited to the use of weapons as means of warfare.

This study takes a step back and critically examines the presupposition that IHL is the primary frame of reference to assess the legality of AWS use. It looks at the use of an AWS in relation to the conduct of hostilities and for law enforcement purposes, both during and outside of an armed conflict. The analysis is based on the premises that during an armed conflict, the use of force by means of an AWS may in some instances fall within a law enforcement and in others within a conduct of hostilities paradigm, and that IHRL applies to the conduct of hostilities concurrently with IHL. The study critically examines how the employment of an AWS may affect the law applicable to the use of force, and explores how the applicability of IHL affects obligations arising under IHRL, thereby affecting the scope for the lawful use of an AWS. A comprehensive treatment of the complex and unsettled interplay between IHRL and IHL is, however, beyond the scope of this paper.

The study analyses international legal instruments and scholarly writings, and draws on human rights jurisprudence, especially case law of the European Court of Human Rights (ECtHR). Although case law on automated killing is relatively rare, a number of human rights cases provide insights into the challenges that algorithm-based decision making generally, and use of force measures specifically, pose to human dignity and the protection of human rights.

The use of autonomous technologies to secure a perimeter or boundary and deny access to and defend a delimited zone, for example, around a military camp, a deten-

22 C. Heyns, 'Human Rights and the Use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement', 38 *Human Rights Quarterly* 2 (2016) 350–378; Amnesty International, *Autonomous Weapons Systems: Five Key Human Rights Issues for Consideration*, 2015, https://www.amnesty.org/en/documents/act30/1401/2015/en/; Human Rights Watch (HRW) and the International Human Rights Clinic at Harvard Law School (IHRC), *Shaking the Foundations: The Human Rights Implications of Killer Robots*, May 2014, https://www.hrw.org/sites/default/files/reports/arms0514_ForUpload_0.pdf.

tion centre or along an international border (so-called 'sentry-AWS') provides the backdrop to the legal discussion. Concentrating on a concrete application of AWS has the advantage of situating technologies whose characteristics are unknown and whose existence is uncertain, within a knowable and regulated context.[23] Whereas a 'hunter-killer scenario', where an AWS with mobile weapon platforms could be used to administer violence in a spatially unbounded manner, tends to be regarded as undesirable by most commentators, the use of an autonomous system to secure a boundary tends to be portrayed by proponents of AWS as a defensive and limited and therefore an *a priori* acceptable application.[24] Looking at AWS as border technology deployed to control (movement, in particular) is interesting not only because this brings into question how human beings exercise control, but also because borders are zones of contestation where the juridical divides between war and peace and between hostilities and law enforcement are manifested.

The study is organized as follows: it starts out by discussing ways of envisioning AWS, human agency and control, and the application of legal rules in the use of AWS. This discussion also serves to circumscribe practices and technologies of concern to the study. Following this, the legal framework governing the use of force is presented. As the applicability of a legal rule is a prerequisite for its application in a concrete situation, the study examines how changes in human intent and control manifested when an AWS is used impact on the applicability of IHRL and IHL standards on the use of force in the context of three unsettled legal debates: extraterritorial obligations under human rights treaties, the threshold for triggering an international armed conflict (IAC) and the belligerent nexus of an act of violence to an ongoing armed conflict. Against the backdrop of jurisprudence on automated killing, the last part of the study investigates IHRL and IHL requirements and constraints on the use of sentry-AWS. The focus is on challenges to the rights to life, freedom of movement, liberty and security of person, privacy, the right not to be discriminated against and not to be subjected to cruel, inhuman or degrading treatment, and the right to an effective remedy. Drawing on critical appraisals of present security practices, the final part of the study elucidates how the process of autonomous, algorithmic target construction threatens human rights and human dignity.

23 On the politics of treating certain aspects as (un)known or (un)knowable in juridical and techno-scientific discourses, see J. Weber, 'Black-Boxing Organisms, Exploiting the Unpredictable: Control Paradigms in Human-Machine Translations', in M. Carrier and A. Nordmann (eds), *Science in the Context of Application*, Springer Science + Business Media B. V., 2011, pp 409–429; E. Datteri and G. Tamburrini, 'Robotic Weapons and Democratic Decision-Making', in E. Hilgendorf and J. P. Günther (eds), *Robotik und Gesetzgebung*, Nomos Verlagsgesellschaft, 2013, 211–230. See also F. Johns, *Non-Legality in International Law: Unruly Law*, Cambridge University Press, 2013, pp 1–14.

24 A. M. Johnson and S. Axinn, 'The Morality of Autonomous Robots', 12 *Journal of Military Ethics* 2 (2013) 137–138, http://www.tandfonline.com/doi/abs/10.1080/15027570.2013.818399.

3. ENVISIONING AUTONOMOUS WEAPON SYSTEMS

There is no common understanding of what an AWS is and at this stage it may not be constructive to tightly define the term. The lack of a common understanding is not simply due to a failure to agree on the correct nomenclature.

Rather, participants in multilateral policy discussions frame issues of autonomy, agency and weapons in different ways. This affects what technologies or practices they identify as problematic and their orientation toward a potential regulatory response.[25] Whereas some seek to exclude existing weapon systems from policy discussions on AWS, others stress that past and present violent practices involving mines, torpedoes, sentry guns, automated anti-missile systems, armed drones and other (highly) automated technologies offer important insights into the changing modes and locales of human agency in the use of force and should be part of the debate.[26] Some consider that the differentiation between automated and autonomous systems will be critical for the debate about future AWS.[27] Others operate with taxonomies describing degrees of autonomy in weapon systems.[28] Another common approach is to focus on the role of human agents in the military decision-making cycle and to distinguish among weapon systems based on whether they have a (hu)man in, on or out of 'the loop'.[29] 'Man-out-of-the-loop' and certain 'on-the-loop' systems are sometimes termed 'fully autonomous' weapons.[30] Irre-

25 For a brief discussion of proposed definitions of AWS within the CCW context, see H. M. Roff, 'Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons', Briefing Paper for Delegates at the Review Conference of the Convention on Certain Conventional Weapons (CCW), Geneva, 12–16 December 2016, Global Security Initiative, Arizona State University, https://globalsecurity.asu.edu/sites/default/files/files/Control-or-Judgment-Understanding-the-Scope.pdf.

26 US DoD, *Law of War Manual*, supra fn 9, s 6.5.9.1, p 329, describes mines as 'rudimentary autonomous weapons'. Drawing on experience with existing technologies to inform the debate on AWS does not mean that such technologies would or should fall within the ambit of a potential future legal restriction on AWS.

27 M. L. Cummings, 'The Human Role in Autonomous Weapon Design and Deployment', [undated], p 5, https://www.law.upenn.edu/live/files/3884-cummings-the-human-role-in-autonomous-weapons.

28 See, e.g., M. Dickow, A. Dahlmann, C. Alwardt, F. Sauer and N. Schörnig, *First Steps Towards a Multidimensional Autonomy Risk Assessment (MARA) in Weapons Systems*, Working Paper no 20, Institute for Peace Research and Security Policy at the University of Hamburg, December 2015, https://ifsh.de/file-IFAR/pdf_deutsch/IFAR-WP20.pdf (proposing a framework to quantify and compute key descriptive characteristics of systems to gauge their autonomous capabilities). See also US DoD, Directive no 3000.09, supra fn 6, Glossary, Part II (distinguishing between autonomous, semi-autonomous, and human-supervised autonomous weapon systems).

29 W. C. Marra and S. K. McNeil, *Understanding "The Loop": Humans and the Next Drone Generations*, Issues in Governance Studies no 50, The Brookings Institution, 2012, https://www.brookings.edu/wp-content/uploads/2016/06/27-humans-drones-marra-mcneil.pdf.

30 HRW, *Losing Humanity: The Case Against Killer Robots*, 2012, https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

spective of whether a weapon system is described as (fully) autonomous or not, the International Committee of the Red Cross (ICRC) has proposed to define an AWS as a weapon system with autonomy in 'critical functions'.[31] Such a system would be able to detect, track, select and attack (e.g. fire at) a target without direct, in the sense of spatially, temporally or causally proximate, human intervention. Finally, an approach centering on the question of whether an AWS operates outside of meaningful human control is gaining increasing traction.[32] Proponents of this approach tend to understand autonomous agency as relational, rather than being located in either the human or the machine.

More fundamentally, perhaps, the lack of agreement on what is problematic about AWS reflects that participants in the debate draw on and generate different socio-technical imaginaries about how military technology evolves and what role society does or ought to play in shaping that evolution.[33] Some commentators embrace the narrative that increasing autonomy in weapon systems enables conducting war in ever more moral and legal ways.[34] Among this group, the advent of AWS tends to be portrayed as inevitable, and their preventive prohibition as futile. Others, including the author of this study, situate AWS within a broader trend of automated killing and problematize the expansion of the spatial and temporal dimensions of militarized rationalities and technologies into civilian spheres. These commentators tend to challenge claims about the inevitability of AWS and see a preventive prohibition as a way of shaping technological developments.[35]

For the purposes of this paper (and side-stepping complicated discussions about autonomy and AI), it is sufficient to describe an AWS schematically as a weapon system with sensors, algorithms and effectors.[36] Such a system can include stationary as well as mobile robotic components (e.g. unmanned air, ground or naval vehicles) equipped with active or passive sensors to navigate and detect objects, motion or patterns.[37]

31 ICRC, *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, Report, Expert Meeting, Geneva, Switzerland, 26–28 March 2014, November 2014, p 62.

32 E.g., Article 36, *Structuring Debate on Autonomous Weapons Systems*, supra fn 6.

33 An imaginary can be described as a shared vision (an imagined future), symbols and associated feelings about something. Imaginaries help produce systems of meaning in a society. Scientists, policy makers and other actors draw on and generate imaginaries to inform and justify their actions, thereby shaping scientific and policy developments. Socio-technical imaginaries encode visions of what is attainable through science and technology, as well as about what ought to be attained (see, e.g., S. Jasanoff, 'Future Imperfect: Science, Technology, and the Imaginations of Modernity', in S. Jasanoff and S-H. Kim (eds), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, The University of Chicago Press, 2015, p 4).

34 Anderson and Waxman, *Law and Ethics for Autonomous Weapon Systems*, supra fn 20, p 2.

35 E.g. N. E. Sharkey, 'The Evitability of Autonomous Robot Warfare', 94 *International Review of the Red Cross* (*IRRC*) 886 (2012) 787–799.

36 Defense Science Board (DSB), *Summer Study on Autonomy*, Report, US DoD, June 2016, p 11, conceptualizes 'technologies critical to the development of autonomous systems' in terms of 'sense, think/decide, act, team', http://www.acq.osd.mil/dsb/reports/2010s/DSBSS15.pdf.

37 Sensors can include electro-optical, infrared, radar or sonar. For a recent survey of sensors used to detect persons, see, e.g., 'Detectors of Humans', in J. Fraden, *Handbook of Modern Sensors: Physics, Designs, and Applications*, Springer International Publishing, 2016, pp 271–333, doi: 10.1007/978-3-319-19303-8_7.

The collected sensor data is processed by computationally intensive algorithms enabling the detection, tracking and classification of objects. Targets can be identified by comparing sensor data with target types contained in a database or perception library.[38] Finally, the system includes a weapon or munition to 'engage' a target, and some sort of communication system that allows for human interaction.

Neither autonomous cars nor remote-controlled, unmanned, weapon platforms (armed drones) that can navigate or run diagnostics without human intervention, autonomously select targets of attack. Nevertheless, differentiating an AWS from a 'non-weaponized' autonomous system is not straightforward.[39] It is disputable what technologies are, what they are for and how they are implicated in the use of force.[40] Technologies and practices of violence are mutually constitutive and shape each other. In the absence of an internationally agreed definition of a weapon (system), weapons are commonly described in legal practice as devices that by *design*, *use* or *intended use* are capable of causing incapacitation, injury, illness, severe mental suffering or death of persons, damage to or destruction of objects or loss of functionality.[41] A weapon, together with other devices, materials, instruments, mechanisms, equipment or software, can form a weapon system.[42] In the context of hostile activities in the cyber domain where similar definitional challenges arise, Harrison Dinniss stresses that 'the key factors that determine its use as a weapon is not the nature of the object itself, but rather how the object was used, against whom and why'.[43] Given the tightening connection between surveillance

38 W. H. Chun and N. Papanikolopoulos, 'Robot Surveillance and Security', in B. Siciliano and O. Khatib (eds), *Springer Handbook of Robotics*, 2nd edn, Springer International Publishing, 2016, p 1613.

39 Considertheheadline'FBISaysAutonomousVehiclesCouldBeLethalWeapons',*GTMagazine*,17July2014, http://www.govtech.com/transportation/FBI-Says-Autonomous-Vehicles-Could-Be-Lethal-Weapons.html. According to a speaker at an ICRC expert meeting, Platforma-M, a Russian system reportedly under development, is 'designed to carry out rescue missions' but 'could also be used to lay smoke screens and plant mines' (ICRC, *Autonomous Weapon Systems*, Expert Meeting Report (2016), supra fn 6, p 21).

40 Critiquing the doctrine of technological neutrality in the context of AWS, I. Kerr and K. Szilagyi, 'Evitable Conflicts, Inevitable Technologies? The Science and Fiction of Robotic Warfare and IHL', *Law Culture and the Humanities* (online 7 January 2014) 18–27, doi: 10.1177/1743872113509443. Generally, B. Rappert, *Controlling the Weapons of War: Politics, Persuasion, and the Prohibition of Inhumanity*, Routledge, 2006.

41 See, e.g., M. Roscini, *Cyber Operations and the Use of Force in International Law*, Oxford University Press, 2014, pp 168–169.

42 Program on Humanitarian Policy and Conflict Research at Harvard University (HPCR), *Manual on International Law Applicable to Air and Missile Warfare* (AMW Manual), 2009, §A(1)(ff), http://ihlresearch.org/amw/HPCR%20Manual.pdf; HPCR, *Commentary on the Manual on International Law Applicable to Air and Missile Warfare* (Commentary on AMW Manual), 2010, p 55, http://ihlresearch.org/amw/Commentary%20on%20the%20HPCR%20Manual.pdf; M. Schmitt (ed), *Tallinn Manual on the International Law Applicable to Cyber Warfare*, Cambridge University Press, 2013, pp 141–142, https://issuu.com/nato_ccd_coe/docs/tallinnmanual.

43 H. Harrison Dinniss, *Cyber Warfare and the Laws of War*, Cambridge University Press, 2012, p 70. See also Anderson et al, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems', supra fn 19, 397 (recognizing that whether a system operates autonomously is, inter alia, a function of how it is operated and controlled and of the operational context and conditions).

and targeting technologies and practices,[44] and considering that some AWS components are intangible and can be geographically distributed, it is far from clear when their use or intended use constitutes an AWS, that is, where and when an AWS begins and ends. This presents a challenge to the construction of AWS as a regulatory category in the framework of the CCW, where deliberations tend to focus on the materialities of weapon control.

TECHNOLOGIES FOR AUTONOMOUS AREA DENIAL, BORDER CONTROL AND PERIMETER SECURITY

One of the functions envisaged for AWS is to prevent people or vehicles from crossing a line or entering or exiting an area. Sentry systems are being advertised to survey and guard boundaries and patrol areas. Partisans of such technologies expect that 'the careful insertion of automatic and autonomous technologies'[45] will obviate the need to dispatch human security personnel to respond to 'emerging incidents', 'prob[e] maneuvers by enemy squads' or 'intercept intruders'.[46] They also hope that the reduced need for a permanent physical presence of human guards will reduce manpower requirements.[47] The other goal is to 'close the kill chain' by combining target detection, identification and the capability to fire.[48]

Sentry systems with autonomous capabilities in critical functions are already deployed, but none of them is currently selecting and attacking targets without direct human intervention. One system of this type is DoDaam's Super aEgis II deployed in the Demilitarized Zone (DMZ) between North and South Korea.[49] The system is advertised as being able to detect humans from 3 kilometers away in daylight and from 2.2 kilometers at night. According to the manufacturer, it can be equipped, among other options, with a 12.7mm machine gun, a 40mm grenade launcher or a surface-to-air missile launcher.

A human operator can specify the perimeter within which the system scans for targets,[50] and the system reportedly has the capability to 'identify, track and destroy a moving target' and to issue a warning to a target before an attack. The original version 'had an auto-firing system', enabling it to target and attack 'without human intervention', but in present practice a human operator unlocks the system's firing ability.[51]

Another sentry system is deployed by Israel along its border with Gaza. The 'Roeh-Yoreh' ('Sees-Fires') Sentry Tech system comprises remotely operated, pre-positioned sensor-to-shooter weapon platforms.[52] These are equipped with Rafael's Mini-Samson weapon station, mounting a machine gun of 5.56 or 7.62mm calibre. The weapon station can also carry machine guns of a larger calibre or a 40mm grenade launcher.[53] A newer variant can deliver long-range anti-tank guided missiles, enabling 'strikes on distant targets'.[54] According to one source 'The idea, ultimately, is to have a "closed-loop" system – no human intervention required'.[55] In present practice, however, a human operator pulls the trigger and the 'operator cannot engage a sensor-acquired target without verification through the weapon station [electro-optical] package'.[56]

Sentry systems can comprise mobile units that allow them to patrol specific areas or perimeters. These robots can follow patrol paths determined in advance by human operators, or swarms of mobile units can be left to self-organize within a predetermined area. In the latter case, patrolling paths emerge based on the robots' interactions with the environment.[57] The Guardium, an unmanned ground vehicle developed by G-Nius follows pre-programmed routes. The unit is said to be able to 'navigate alone through cities' or 'patrol borders'.[58] It is currently deployed by Israel along its borders with Gaza and Lebanon and has previously been used at Ben Gurion Airport. The Guardium carries various sensors, including video and thermal cameras, and is equipped 'with auto-target acquisition and capture'.

44 K. H. Kindervater, 'The Emergence of Lethal Surveillance: Watching and Killing in the History of Drone Technology', 47(3) *Security Dialogue* (2016) 224 (describing 'lethal surveillance' as a practice where intelligence, surveillance and reconnaissance capabilities 'are linked directly to targeted killing in an attempt to close the temporal and spatial gap between the two … a practice in which mechanisms of surveillance and knowledge production and decisions on life and death have become one and the same.' See also 'Surveillance and Annihilation', in Chamayou, *Drone Theory*, supra fn 1, pp 37–45; L. Amoore, 'Algorithmic War: Everyday Geographies of the War on Terror', 41 *Antipode* 1 (2009) 49–69; T. Wall and T. Monahan, 'Surveillance and Violence from Afar: The Politics of Drones and Liminal Security-Scapes', 15 *Theoretical Criminology* 3 (2011) 239–245.

45 Chun and Papanikolopoulos, 'Robot Surveillance and Security', supra fn 38, p 1606.

46 N. Shachtman, 'Robo-Snipers, "Auto Kill Zones" to Protect Israeli Borders', *Wired*, 6 April 2007, https://www.wired.com/2007/06/for_years_and_y/; B. Shoop, M. Johnston, R. Goehring, J. Moneyhun and B. Skibba, *Mobile Detection Assessment and Response Systems (MDARS): A Force Protection, Physical Security Operational Success*, Space and Naval Warfare Systems Center, San Diego, http://www.dtic.mil/dtic/tr/fulltext/u2/a449408.pdf.

47 Shoop et al, *Mobile Detection Assessment and Response Systems (MDARS)*, supra fn 46.

48 Defense Update, 'Lethal Presence: Remotely Controlled Sentries Assume Guard Roles', 27 November 2008, http://defense-update.com/20081127_sentrytech.html.

49 DoDaam Systems Ltd, 'Super aEgis II', http://www.dodaam.com/eng/sub2/menu2_1_4.php.

50 S. Parkin, 'Killer Robots: The Soldiers that Never Sleep', *BBC Future*, 16 July 2015, http://www.bbc.com/future/story/20150715-killer-robots-the-soldiers-that-never-sleep.

51 A step taken due to clients' concerns. For DoDaam engineers, however, the requirement of human intervention is 'a temporary state'. 'Their aim is now to make the product "smarter" by focusing on "increasing the gun's automatic functionality"' (Parkin, 'Killer Robots', supra fn 50).

52 R. Hughes, 'IDF Deploys Sentry Tech on Gaza Border', *Jane's Defence Weekly*, 6 June 2007.

53 RAFAEL Advanced Defence Systems Ltd, 'Samson Mini RWS Compact Stabilized Remote Weapon Station', http://www.rafael.co.il/5700-744-en/Marketing.aspx.

54 Hughes, 'IDF Deploys Sentry Tech on Gaza Border', supra fn 52.

55 Shachtman, 'Robo-Snipers', supra fn 46.

56 Hughes, 'IDF Deploys Sentry Tech on Gaza Border', supra fn 52.

57 F. Legras, A. Glad, O. Simonin and F. Charpillet, 'Authority Sharing in a Swarm of UAVs: Simulation and Experiments with Operators', in S. Carpin, I. Noda, E. Pagello, M. Reggiani and O. von Stryk (eds), *Simulation, Modeling, and Programming for Autonomous Robots*, Springer, 2008. https://link.springer.com/chapter/10.1007%2F978-3-540-89076-8_29#page-1.

58 Associated Press, 'Israeli Military Unveils Armed Patrol Robot', *Fox News*, 28 April 2008, http://www.foxnews.com/story/2008/04/28/israeli-military-unveils-armed-patrol-robot.html.

It can be equipped with a variety of remotely controlled 'lethal or less than lethal weapons'.[59] Similarly, GDSR's Mobile Detection Assessment and Response System (MDARS) provides 'automated intrusion detection' in US Department of Defense warehouses and storage sites as well as 'nuclear sites'.[60] The robotic platform is capable of autonomous movement within 'a defined area of operation' or 'an enclosed security area whose boundaries are pre-programmed. It includes motion detection and incident assessment subsystems.[61] The MDARS is fitted with an 'operator-controlled', 'non-lethal gun pod'.[62]

Autonomous patrol and sentry systems are advertised for use in diverse operational environments. They appear to be in high demand in 'the burgeoning homeland security industries around the globe'.[63] The Super aEgis II is reportedly 'in active use in numerous locations in the Middle East, including three airbases in the United Arab Emirates ... the Royal Palace in Abu Dhabi, an armoury in Qatar and numerous other unspecified airports, power plants, pipelines and military airbases elsewhere in the world'.[64] Sentry Tech exists as a mobile station that is 'easily transportable and can, for example, be deployed to protect temporary forward/base camps in expeditionary/peacekeeping operations'.[65] The MDARS is advertised for 'random patrols around inventory-sensitive warehouses, air-fields, ammunition supply depots, and port facilities', as well as to support 'force protection efforts in the battlespace or for homeland security and border patrol efforts across the US and its territories'.[66] Whereas the MDARS was initially conceived for 'structured/semi-structured facilities', follow-on projects aim to expand the scope of application into the 'tactical unstructured environment'. The US Army's Family of Integrated Rapid Response Equipment (FIRRE) is intended for operations 'outside a defined perimeter on semi to unstructured terrain in support of force protection/physical security missions in a more hostile environment'.[67]

59 'Enguard! Introducing the Guardium UGV', *Defense Update*, https://defense-update.com/products/g/guardium.htm.

60 SPAWAR Systems Center Pacific, 'Mobile Detection Assessment and Response System (MDARS)', http://www.public.navy.mil/spawar/Pacific/Robotics/Pages/MDARS.aspx. The Russian Strategic Missile Forces have reportedly announced 'that mobile robots would be standing guard over five ballistic missile installations' (D. Hambling, 'Armed Russian Robocops to Defend Missile Bases', *New Scientist*, 23 April 2014, https://www.newscientist.com/article/mg22229664-400-armed-russian-robocops-to-defend-missile-bases/).

61 Shoop et al, *Mobile Detection Assessment and Response Systems (MDARS)*, supra fn 46, 3.

62 General Dynamics Robotic Systems, 'MDARS' Brochure, TechyLib, https://www.techylib.com/en/view/pillowfists/mdars_general_dynamics_robotic_systems

63 J. Cook, 'Israel's Video Game Killing Technology', *The Electronic Intifada*, 13 July 2010, https://electronicintifada.net/content/israels-video-game-killing-technology/8919.

64 Parkin, 'Killer Robots', supra fn 50.

65 Hughes, 'IDF Deploys Sentry Tech on Gaza Border', supra fn 52.

66 General Dynamics Robotic Systems, 'MDARS', supra fn 62.

67 Shoop et al., *Mobile Detection Assessment and Response Systems (MDARS)*, supra fn 46, 3.

A. HUMAN CONTROL AND THE USE OF FORCE

As advertisements for sentry-AWS illustrate, cultures of control and a belief in the controllability of unknown future threats are important drivers of algorithm-based security practices and technologies.[68] Human beings have long used technologies to exercise control over the natural and human world, as well as to deploy violence. That weapons and their consequences are controlled and controllable is a long-standing requirement for the moral acceptability, political legitimacy and legality of organized violence. The latter is, for instance, reflected in the IHL requirement that the harmful effects of weapons must not be unforeseeable or escape, either in space or in time, the control of those who employ them,[69] as well as in IHRL standards on the use of force demanding that state agents 'place the flow of events under their control' in law enforcement operations.[70] The exercise of control by state agents is a legal requirement and, at the same time, it plays a role in delimiting the boundaries of state responsibility under international law for the consequences of armed violence.[71]

Changes in how human beings exercise control in the use of weapons affect their ability and, by extension, that of the state on whose behalf they act, to perform legal duties and be accountable for the consequences. For one, tasking a machine with selecting and firing at targets makes it more difficult for the user to predict every *particular* target, the *precise* moment and location where violence is administered and the *concrete* environment within which violent effects are produced. Users of an AWS are in principle unable to predict and control completely its behaviour. If an AWS functions on the basis of a model of the environment within which it operates, any unforeseen change to that environment, or operation outside of it, can lead to unpredictability in its functioning.[72] As Suchman and Weber explain, 'plans and any other form of prescriptive specification presuppose competencies

68 DSB, *Summer Study on Autonomy*, supra fn 36, pp 80–81 (envisaging the development of an autonomous system that would 'sense the state of the world and build an internal representation of the underlying causal linkages' so as to predict 'geopolitical events', with a view to 'safeguard U.S. interests').

69 Art 14, 1956 ICRC Draft Rules for the Limitation of the Dangers incurred by the Civilian Population in Time of War; International Court of Justice (ICJ), *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 8 July 1996, §35. For more sources, see 'Practice Relating to Rule 71 – Weapons That Are by Nature Indiscriminate', ICRC, Customary IHL Database (ICRC CIHL Database), https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule71.

70 European Court of Human Rights (ECtHR), *Mikayil Mammadov v Azerbaijan*, App no 4762/05, Judgment, 17 December 2009, §114.

71 This is reflected, for example, in provisions on the attribution of conduct to a state and in circumstances precluding wrongfulness. See, e.g., Arts 8 and 23(1), 2001 International Law Commission Articles on the Responsibility of States for Internationally Wrongful Acts (UNGA Res 56/83, 12 December 2001).

72 ICRC, *Autonomous Weapon* Systems, Expert Meeting Report (2016), supra fn 6, p 8. P. Lin, G. Bekey and K. Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, Ethics + Emerging Sciences Group at California Polytechnic State University, San Luis Obispo, December 2008, p 8, http://ethics.calpoly.edu/ONR_report.pdf (describing the 'common misconception that robots will do only what we have programmed them to do' as a 'sorely outdated' belief, given the complexity of programs and potentially emergent behaviours).

and in situ forms of interaction that they can never fully specify'.[73] The challenge is compounded if complex distributed systems, based on swarm intelligence for example, are used where collective behaviour can *emerge* from the self-organized interactions of system components with each other and their environment,[74] or if AWS are made to learn so as to function in dynamic, unstructured environments (the 'real world'). How machines 'make decisions' and learn is not well understood today,[75] and the underlying premise that a representation of our world can be adequately encoded so as to ensure that the consequences of AWS use comply with legal precepts is deeply contentious.[76] This is a major concern, not least for those who risk being adversely affected by algorithm-based decisions. At what point human control is no longer exercised in a meaningful or appropriate manner is, thus, a key question in the debate on AWS.[77]

In addition to constraints based on ethical and other imperatives, compliance with the law presupposes a measure of human agency in the use of force that places limitations on permissible 'human-machine configurations'.[78] Legal obligations are addressed to human beings.[79] Although the use of anthropomorphizing language can be misleading,[80] this author espouses the view that an AWS does not make legal judgements (as opposed to algorithmic calculations).[81] From this anthropocentric perspective, an AWS is an artefact – an object made for a certain purpose, devoid of intentionality.[82] It cannot meaningfully be treated as a holder of rights or an entity accountable for harm done or infringements of the law.[83]

73 L. Suchman and J. Weber, 'Human-Machine Autonomies', in N. Bhuta, S. Beck, R. Geiss, H.-Y. Liu and C. Kress (eds), *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge University Press, 2016, p 85. DSB, *Summer Study on Autonomy*, supra fn 36, p 18 (recognizing the potential for 'surprise during operations' as 'many autonomous system behaviors will change over time due to learning', leading to discrepancies between actual system performance and operator expectations).

74 M. Dorigo and M. Birattari, 'Swarm Intelligence', 2 *Scholarpedia* 9 (2007) 1462, doi: 10.4249/scholarpedia.1462; P. Scharre, *Robotics on the Battlefield Part II: The Coming Swarm*, CNAS, 2014, https://s3.amazonaws.com/files.cnas.org/documents/CNAS_TheComingSwarm_Scharre.pdf; DSB, *Summer Study on Autonomy*, supra fn 36, pp 83–87 (noting that 'the hundreds-to-thousands of individual platforms would be beyond the ability of humans to control directly' (p 86)). See also Weber, 'Black-Boxing Organisms', supra fn 23, p 423 (explaining how the vision of emergent behaviour has become part of the leitmotif of a new techno-rationality that strives for AI systems that operate autonomously in open and complex environments).

75 B. Goodman and S. Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"', Paper presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, 28 June 2016, https://arxiv.org/abs/1606.08813; J. Pearson, 'When AI Goes Wrong, We Won't Be Able to Ask It Why', *Motherboard*, 6 July 2016, https://motherboard.vice.com/read/ai-deep-learning-ethics-right-to-explanation.

76 Suchman and Weber, 'Human-Machine Autonomies', supra fn 73, pp 85–86. See also N. Soares, 'The Value Learning Problem', Machine Intelligence Research Institute, Technical Report no 2015-4, https://intelligence.org/files/obsolete/ValueLearningProblem.pdf ('Human goals are complex, culturally laden, and context-dependent', p 1).

77 See, e.g., 'The Concept of "Meaningful Human Control"', Working Paper submitted by Austria, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 13–17 April 2015, http://bit.ly/1Nx8eLb.

78 Suchman and Weber, 'Human-Machine Autonomies', supra fn 73, p 78 (arguing that 'contemporary social theory has effectively challenged the premise that autonomy can be adequately understood as being an intrinsic capacity of an entity, whether human or machine, shifting the focus instead to the capacities for action that arise out of particular socio-technical systems.' The concept of 'configuration' draws attention to the relations between human beings and machines, supporting an understanding of autonomous agency as relational).

79 See, in particular, US DoD, *Law of War Manual*, supra fn 9, s 6.5.9.3, p 330, according to which 'The law of war rules on conducting attacks … impose obligations on persons. These rules do not impose obligations on the weapons themselves; of course, an inanimate object could not assume an "obligation" in any event.' See also, 'Towards a "Compliance-Based" Approach to LAWS', supra fn 10, p 3 , §16 (noting that 'a manifest presumption of human agency' is reflected in a range of IHL provisions); Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, fn 53, 362 (stating that 'an unspoken assumption' of IHRL is that 'the decision to use lethal force must be reasonable and taken by a human').

80 On the pitfalls of anthropomorphization in the context of AWS, see N. Sharkey and L. Suchman, 'Wishful Mnemonics and Autonomous Killing Machines', 136 *AISB Quarterly* (2013) 14–22; K. Zawieska, 'Do Robots Equal Humans? Anthropomorphic Terminology in LAWS', Presentation, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 13–17 April 2015, http://www.unog.ch/80256EDD006B8954/(httpAssets)/369A75B470A5A368C1257E290041E20B/$file/23+Karolina+Zawieska+SS.pdf.

81 See, e.g., E. Lieblich and E. Benvenisti, 'The Obligation to Exercise Discretion: Why Autonomous Weapons Systems are Unlawful', in N. Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge University Press, 2016, pp 252–253 (pointing out that while it is possible and acceptable to many that computers apply rules (e.g. calculating an artillery projectile trajectory), one is 'hard-pressed to imagine them applying standards, which per se do not aim to predict the right legal outcome in any given situation', as is required for determining direct participation in hostilities or *jus in bello* proportionality, for instance). Suggestions to the effect that an AWS would apply the law are frequently, although perhaps sometimes unwittingly, made. Consider, e.g., Advisory Council on International Affairs and the Advisory Committee on Issues of Public International Law, *Autonomous Weapon Systems: The Need for Meaningful Human Control*, no 97 AIV / no 26 CAVV, October 2015, p 26, asserting that autonomous weapons 'will not be able to independently apply IHL for at least the next 10 years'. The statement implies that complex enough algorithmic calculations can be equivalent to the 'application of law'. This orientation has to be viewed with scepticism if applying law is understood as an interpretive act involving the construction of a socially acceptable meaning of an indeterminate legal norm in application to specific facts, and if it is recognized that fact and law are socially constructed and that any fact or reality has a normative origin that gives it meaning (see, e.g., P. Nerhot (ed), *Law, Interpretation and Reality: Essays in Epistemology, Hermeneutics and Jurisprudence*, Springer Science + Business Media Dordrecht, 1990, p 2). The discussion on the locus of agency intersects with the 'codifiability debate'. For a brief overview, see D. Purves, R. Jenkins and B. J. Strawser, 'Autonomous Machines, Moral Judgment, and Acting for the Right Reasons', 18 *Ethical Theory and Moral Practice* 4 (2015) 851–872, doi: 10.1007/s10677-015-9563-y (arguing that 'even a sophisticated robot is not the kind of thing that is capable of replicating human moral judgment' on the basis that 'human moral judgment is not codifiable, i.e. it cannot be captured by a list of rules', and that even if robot 'decisions' are extensionally indistinguishable from human moral judgment *in their result*, they 'could not be made for the right reasons', rendering them 'morally deficient'). For a different view, see R. C. Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, Technical Report GIT-GVU-07-11, Georgia Institute of Technology, 2007, p 7, http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf ('I am convinced that they can perform more ethically than human soldiers are capable of').

82 On the ascription of moral agency to machines, see P. M. Asaro, 'Determinism, Machine Agency, and Responsibility', 2 *Politica & Società* (2014) 265–292, doi: 10.4476/77103.

83 In contrast, see Draft Report, European Parliament, Committee on Legal Affairs, 2015/2103(INL), 31 May 2016, p 12, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-582.443+01+DOC+PDF+V0//EN (contemplating electronic personhood for 'at least the most sophisticated autonomous robots'). On the 'status question', see S. Beck, 'Über Sinn und Unsinn von Statusfragen – zu Vor- und Nachteilen der Einführung einer elektronischen Person', in E. Hilgendorf and J. P. Günther (eds), *Robotik und Gesetzgebung*, Nomos Verlagsgesellschaft, 2013, 239–260.

Legal norms are one way to formally limit human–machine configurations in order to ensure that human beings retain meaningful control in the use of force.[84] In present practice, human control takes the form of technical and normative restrictions and requirements pertaining to the reasons why and the manner in which force is used, when and where force is applied or violent effects are produced, and who or what is harmed – both in respect of persons and objects that force is directed at and that may be incidentally affected.[85] For example, to retain a measure of control over the violent effects of landmines, States Parties to the 1996 CCW Amended Protocol II accept responsibility for recording the location of mines,fencing and perimeter-marking mined areas, rendering mines inoperable through technical measures, and clearing them after a specified lapse of time.[86]

This example illustrates that human control over weapon effects does not need to be absolute. Today, it is accepted that weapons produce effects 'on their own' within specified spatio-temporal boundaries and according to predefined parameters.[87] The example of landmines also demonstrates, however, that where these boundaries should be drawn can be controversial and can change over time. A number of states have concluded that the adverse impact of anti-personnel landmines on human lives and livelihoods cannot be adequately controlled through the Protocol's spatio-temporal restrictions and procedural requirements – a normative development formalized in the comprehensive legal ban on anti-personnel mines adopted in 1997.[88]

84 As reflected in African Commission on Human and Peoples' Rights (AfCommHPR), *General Comment no 3 on the African Charter on Human and Peoples' Rights: The Right to Life (Article 4)*, 2015, s F, §35: 'Any machine autonomy in the selection of human targets or the use of force should be subject to meaningful human control. The use of such new technologies should follow the established rules of international law.'

85 Aspects of how human control is exercised over weapons can be conceptualized in terms of 'proxy indicators' and 'space-time partitions' (Article 36, *Structuring Debate on Autonomous Weapons Systems*, supra fn 6), 'dynamic diligence' (P. Margulies, 'Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts', in J. Ohlin (ed), *Research Handbook on Remote Warfare*, Edward Elgar Press, forthcoming. Roger Williams University School of Law, Legal Studies Research Paper 166, available at SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2734900), 'levels of human supervisory control' (N. Sharkey, 'Towards a Principle for the Human Supervisory Control of Robot Weapons', 2 *Politica & Società* (2014) 305–324, doi: 10.4476/77105), or perhaps even 'multidimensional autonomy risk assessment scores' (M. Dickow et al, *First Steps*, supra fn 28).

86 Arts 3(2), 5(2)(a)-(b) and 10(1) and Technical Annex, 1996 Amended Protocol II to the CCW on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (CCW AmPII).

87 On the evolving spatial dimension of killing with 'remote and autonomously violent devices', see M. Bolton, 'From Minefields to Minespace: An Archeology of the Changing Architecture of Autonomous Killing in US Army Field Manuals on Landmines, Booby Traps and IEDs', 46 *Political Geography* (2015) 41–53.

88 Art 1, 1997 Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction (APMBT).

4. THE APPLICABLE LAW: IHL AND IHRL STANDARDS ON THE USE OF FORCE

The legal debate on AWS has thus far focused on compliance with IHL. Yet, as the examples given above illustrate, autonomous sentry systems are being advertised for use in military combat situations as well as for other activities, including the patrolling of an international border or the securing of a power plant.

If our experience with armed drones is any guide, IHL would be the dominant legal frame of reference for the use of AWS in some situations, whereas their use in other situations would have to be assessed principally against IHRL standards on the use of force.[89] Which set of standards applies in a given situation can be contested and difficult to determine.

Albeit in different ways, both IHL and IHRL aim to safeguard humanity and protect victims of armed violence, including by placing constraints on the use of force. The protection of human dignity is the common aim of IHRL and IHL.[90] Human rights 'derive from the inherent dignity of the human person'[91] and protect everyone from arbitrary deprivation of life, arbitrary interference with the rights to liberty, security and privacy, and from discrimination. IHRL standards on the use of force, notably the 1990 Basic Principles on the Use of Force and Firearms by Law Enforcement Officials (BPUFF), provide the normative framework for the use of force in law enforcement operations, such as the dispersal of a riot, border governance or any other territorial or extraterritorial measure taken by a state to maintain or restore public security, law and order or to otherwise exercise its authority or power over individuals, objects or territory.[92] In the context of the European Convention on Human Rights (ECHR), the use of force must be absolutely necessary (indispensable, unavoidable) and strictly proportionate to the achievement of a legitimate law enforcement aim, such as to defend a person from unlawful violence, to effect a lawful arrest or to prevent the escape of a person lawfully detained, or 'in action

89 The *jus ad bellum* implications of AWS are beyond the scope of this paper. On this topic, see, H. M. Roff, 'Lethal Autonomous Weapons and Jus ad Bellum Proportionality', 47 *Case Western Reserve Journal of International Law* 1 (2015) 37–52, http://scholarlycommons.law.case.edu/jil/vol47/iss1/7; F. Grimal, 'Missile Defence Shields: Automated and Anticipatory Self-Defence?', 19 *Journal of Conflict and Security Law* 2 (2014) 317–339, doi: 10.1093/jcsl/kru001.

90 Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 367.

91 Art 1, 1948 Universal Declaration of Human Rights; Preamble, 1966 International Covenant on Civil and Political Rights (ICCPR).

92 N. Melzer, *Targeted Killing in International Law*, Oxford University Press, 2008, p 90.

lawfully taken for the purpose of quelling a riot or insurrection'.[93] Even in pursuit of these aims, however (and acknowledging that there is debate on this point), potentially lethal force may not be used except as a last resort in order to protect against an imminent (or grave) threat of death (or serious injury).[94]

Human rights protection of life and physical integrity not only entails states having to refrain from the unlawful taking of life, but also having to take positive steps to secure the right to life within their jurisdiction.[95] Among the 'positive obligations' assumed by states are the duty to put in place an appropriate legal and regulatory framework and procedures that strictly control and limit the use of force, including by making the use of potentially lethal force dependent on a 'careful assessment of the surrounding circumstances';[96] to plan, organize and control the general security set-up and specific operations so as to minimize, to the greatest extent possible, recourse to lethal force and incidental loss of life,[97] and, if such force is used, to minimize the hazard it poses to human life (of bystanders and the suspected offender);[98] provide security forces with adequate equipment and weapons to allow for a differentiated use of force;[99] and conduct some form of an effective official investigation when individuals have been killed as a result of the use of force to secure accountability.[100] The failure to fulfill a positive obligation is a human rights violation.

However, the legal protection of human rights, including against deprivation of life, is not absolute. To take account of the difficulties of securing respect for IHRL in time of war or a 'public emergency threatening the life of the nation', states may take measures that derogate from their obligations under human rights treaties in respect of some rights.'[101] Other rights, including the right to life, are not subject to derogation. Yet, deaths resulting from lawful 'acts of war' are not 'arbitrary' deprivations of life.[102] In times of armed conflict, such acts are governed by IHL rules on the conduct of hostilities, which seek to strike a balance between military necessity and considerations or principles of humanity.[103]

IHL prohibits the use of certain means and methods of warfare,[104] as well as direct attacks on civilians and civilian objects, and the launching of indiscriminate or disproportionate attacks, and it requires that all feasible precautions are taken in attack to avoid and, at any rate, minimize civilian harm.[105] Whereas the precise kind and degree of force that may be used in any given attack cannot be determined *ex ante*, 'considerations of humanity require that, within the parameters set by the specific provisions of IHL, no more death, injury, or destruction be caused than is actually necessary for the accomplishment of a legitimate military purpose in the prevailing circumstances'.[106] In this sense, military necessity demands a context-dependent assessment that serves to limit military actions 'from that which positive IHL does not prohibit *in abstracto* to that which is actually required *in concreto*'.[107]

It is widely recognized today that human rights protection does not cease in times of armed conflict. Consequently, IHL and IHRL can apply concurrently and need to be reconciled.[108] In situations of armed conflict, any exercise by states of their authority or power that does *not* amount to the conduct of hostilities, remains governed by law enforcement standards, but IHRL is to be interpreted and applied in a manner that takes account of IHL rules.[109] And even in situations of hostilities, the applicability of IHL does not eclipse states' obligations under IHRL. The latter is flexible enough to take account of practical difficulties that states may encounter

93 Art 2(2), European Convention on Human Rights (ECHR). ECtHR, *McCann et al v The United Kingdom*, App no 18984/91, Judgment, 27 September 1995, §148.

94 Principle 9, 1990 Basic Principles on the Use of Force and Firearms by Law Enforcement Officials (BPUFF). For a brief discussion, see S. Maslen, *Use of Force in Law Enforcement and the Right to Life: The Role of the Human Rights Council*, Geneva Academy of International Humanitarian Law and Human Rights, 2016, pp 11–14. See also AfCommHPR, *General Comment no 3*, supra fn 84, s E, §27) ('the intentional lethal use of force … is prohibited unless it is strictly unavoidable in order to protect life (making it proportionate) and all other means are insufficient to achieve that objective (making it necessary)').

95 For a conceptualization of positive obligations, see S. Krähenmann, 'Positive Obligations in Human Rights Treaties', PhD Thesis no 949, Graduate Institute of International Studies, Geneva, 2012.

96 ECtHR, *Nachova et al v Bulgaria*, App nos 43577/98 and 43579/98, Judgment, 6 July 2005, §96.

97 *McCann et al*, supra fn 93, §194; ECtHR, *Isayeva, Yusupova and Bazayeva v Russia*, App nos 57947/00, 57948/00 and 57949/00, Judgment, 24 February 2005, §171.

98 Principle 5(b), BPUFF; ECtHR, *Ergi v Turkey*, App no 23818/94, Judgment, 28 July 1998, §80.

99 Principle 2, BPUFF; ECtHR, *Güleç v Turkey*, App no 21593/93, Judgment, 27 July 1998, §71.

100 ECtHR, *McKerr v The United Kingdom*, App no 28883/95, 4 May 2001, §111; ECtHR, *Al-Skeini v The United Kingdom*, App no 55721/07, Grand Chamber, Judgment, 7 July 2011, §164.

101 States may only derogate from their obligations to the extent strictly required by the exigencies of the situation, and provided that such measures are not inconsistent with their other obligations under international law and are not discriminatory (Art 4, ICCPR).

102 This is explicit in Art 15(2), ECHR. Another exception are judicial executions based on the death penalty, provided for in, e.g., Art 6(2), ICCPR.

103 On the normative concept of 'humanity' and its relationship to the 'Martens clause', see M. Zagor, *Elementary Considerations of Humanity*, ANU College of Law Research Paper no 12–19, available at SSRN, https://ssrn.com/abstract=2089115. On AWS, the Martens clause and the notion of meaningful human control, see P. M. Asaro, 'Jus Nascendi: Robotic Weapons and the Martens Clause', in R. Calo, A. M. Froomkin and I. Kerr (eds), *Robot Law*, Edward Elgar Publishing, 2016, pp 267–386, doi: 10.4337/9781783476732.00024. On connections between the notions of humanity and human dignity in relation to AWS, see O. Ulgen, 'Human Dignity in an Age of Autonomous Weapons: Are we in Danger of Losing an "Elementary Consideration Of Humanity"?', 2017, European Society of International Law (ESIL) 2016 Annual Conference (Riga), available at SSRN, https://ssrn.com/abstract=2912002.

104 Notably weapons that are by nature indiscriminate or that are of a nature to cause superfluous injury or unnecessary suffering, including weapons that render death inevitable (ICRC CIHL Database, Rules 70 and 71, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul).

105 These rules are of part of customary IHL and apply in both international armed conflicts (IACs) and non-international armed conflicts (NIACs) (ibid, Rules 1, 11, 14 and 15).

106 N. Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law* (DPH Guidance), ICRC, 2009, p 77, https://www.icrc.org/eng/assets/files/other/icrc-002-0990.pdf.

107 Melzer, *Targeted Killing in International Law*, supra fn 92, p 297.

108 C. Droege, 'The Interplay Between International Humanitarian Law and Human Rights Law in Situations of Armed Conflict', 40 *Israel Law Review* 2 (2007) 311.

109 For a discussion on how IHL and IHRL can be applied jointly in a complementary fashion, see G. Gaggioli and R. Kolb, 'A Right to Life in Armed Conflict? The Contribution of the European Court of Human Rights', 37 *Israel Yearbook on Human Rights* (2007) 115–161.

in exercising their authority in such situations so as not to impose an impossible burden on states.[110]

The distinction between law enforcement and conduct of hostilities can have a crucial impact on the humanitarian consequences of an operation since the normative content of these paradigms differs in important respects.[111] The applicability of IHL can change and in some cases significantly diminish the legal protection of life because a conduct of hostilities framework is generally more permissive in its regulation of the use of force than a law enforcement one. However, it is worth keeping in mind that IHRL 'allows the use of lethal force against *anyone* where this is "absolutely necessary" for a legitimate purpose', whereas 'IHL ... prohibits direct attacks against certain categories of persons in absolute terms, that is to say, even in case of "absolute necessity" for a legitimate purpose.'[112]

What legal rules apply to the use of force, and how IHRL and IHL interact, significantly affects the scope for the lawful use of AWS. These questions are also subject to ongoing legal debate. Although this debate cannot be definitely settled here, three controversies are briefly exposed below:

- Does the use of force abroad by means of an AWS establish a sufficient jurisdictional link on its own for the extraterritorial application of IHRL treaties?
- Can an AWS trigger an IAC 'on its own', bringing IHL into operation?
- During an armed conflict, when does the use of force by means of an AWS constitute use of a 'means of warfare', governed by the conduct of hostilities paradigm?

These controversies are instructive for the debate on AWS. First, because issues of control and intent play a pivotal role in these controversies; and second, because if there is no widely shared agreement about what law applies, then this limits our ability to assess the legality of AWS use in light of existing law.

A. HUMAN RIGHTS TREATY OBLIGATIONS ABROAD: AWS AND EXTRA-TERRITORIAL CONTROL

Autonomy in weapon systems allows for increasing distance between the user of an AWS and the place where violence is experienced, including extraterritorially. Armed drones (possible components of an AWS) already allow for the application of lethal force abroad, in places where state agents are not physically present.

110 Ibid, 129 (noting that several aspects of positive obligations 'can be flexibly applied, adapted and developed for situations of armed conflicts').

111 For a discussion of the distinguishing features of and differences between the two 'paradigms', see G. Gaggioli, *The Use of Force in Armed Conflicts: Interplay Between the Conduct of Hostilities and Law Enforcement Paradigms*, Report, Expert Meeting, January 2012, ICRC, November 2013, https://www.icrc.org/eng/assets/files/publications/icrc-002-4171.pdf.

112 Melzer, *Targeted Killing in International Law*, supra fn 92, p 384. Likewise, in certain circumstances, IHL provides stronger protection against the destruction of civilian property than IHRL.

Although the conditions and modalities governing the lawful use of lethal force under customary IHRL – binding on all states – 'virtually coincide with the conventional right to life',[113] ascertaining the extraterritorial applicability of human rights *treaty* law has practical relevance, not least in terms of the availability of remedies to victims in the form of treaty-based human rights mechanisms.[114]

Under human rights treaties, states parties assume obligations to secure to everyone *within their jurisdiction* the rights and freedoms guaranteed by the treaty. It is accepted that the notion of jurisdiction is primarily territorial. To what extent human rights treaties apply when a state performs acts outside of its territory or acts that produce effects there, whether lawfully or unlawfully, is not definitely settled.[115] The trend is toward asserting that states remain bound by at least some of their obligations when they affect individuals abroad.[116] The ECtHR has exceptionally admitted the extraterritorial application of the ECHR in circumstances where a state, through military action, exercises effective overall control of an area outside its national territory,[117] or where the exercise of public powers on the territory of another state brings an individual into the 'physical power and control' or the 'control and authority' of a foreign state (such as when a person is taken into foreign state agents' custody).[118]

113 Melzer, *Targeted Killing in International Law*, supra fn 92, p 211.

114 However, victims of AWS use abroad can be expected to face formidable challenges in accessing the courts of the user state, as well as supra-national human rights mechanisms (HRW and IHRC, *Mind the Gap: The Lack of Accountability for Killer Robots*, April 2015, pp 27–29, https://www.hrw.org/sites/default/files/reports/arms0415_ForUpload_0.pdf).

115 Note that the wording of provisions on the scope of application of human rights treaties differs from one treaty to another, and that some states, the US for example, reject that the ICCPR applies extraterritorially. See, e.g., 'United States Response to the OHCHR Questionnaire on the "Right to Privacy in the Digital Age"', http://www.ohchr.org/Documents/Issues/Privacy/United%20States.pdf. Questions pertaining to jurisdiction can also arise when a state has lost effective control over part of its territory as may happen in connection with an armed conflict or military occupation. In such cases, there is a rebuttable presumption that the territorial state exercises jurisdiction or competence, which bears the burden to show that exceptional circumstances limit its responsibility to respect and ensure respect for human rights (see, e.g., ECtHR, *Sargsyan v Azerbaijan*, App no 40167/06, Grand Chamber, Judgment, 16 June 2015, §§126–131).

116 M. Milanovic, 'Al-Skeini and Al-Jedda in Strasbourg', 23 *European Journal of International Law* (*EJIL*) 1 (2012) 121–139, doi: 10.1093/ejil/chr102; D. Hart, 'War Remains Inside the Court Room: Jurisdiction under ECHR', *UK Human Rights Blog*, 11 September 2016, https://ukhumanrightsblog.com/2016/09/11/war-remains-inside-the-court-room-jurisdiction-under-echr/.

117 ECtHR, *Loizidou v Turkey*, App no 15318/89, Grand Chamber, Judgment, 18 December 1996, §56.

118 *Al-Skeini*, supra fn 100, §§136–137; ECtHR, *Öcalan v Turkey*, App no 46221/99, Grand Chamber, Judgment, 12 May 2005, §91; ECtHR, *Medvedyev et al v France*, App no 3394/03, Grand Chamber, Judgment, 29 March 2010, §67; UN Human Rights Committee (HRCttee), *Sergio Euben Lopez Burgos v Uruguay*, Comm no R.12/52, UN doc supp no 40 (A/36/40) at 176 (1981), §12(1)–(3); Inter-American Commission on Human Rights (IACommHR), *Coard et al v United States*, Report no 109/99, Case 10.951, 29 September 1999, Annual Report 1999, §37.

Controversy persists about whether the use of force abroad constitutes, on its own, a sufficient link to extend jurisdiction extraterritorially.[119] In *Banković et al v Belgium et al*, the ECtHR controversially held that bombardment from the air abroad did not bring the affected people within the jurisdiction of the NATO forces conducting the airstrikes.[120] In contrast, the Court considered in a later case that deaths caused by fire discharged from Turkish helicopters in the Turkey-Iran border area fell within the jurisdiction of Turkey irrespective of the precise location of the victims.[121] In *Al-Skeini v The United Kingdom* the ECtHR determined that the United Kingdom (UK) 'exercised authority and control over individuals killed in the course of … security operations' taking place in an area where the UK exercised the public powers necessary to maintain security.[122] Similarly, in *Jaloud v The Netherlands*, the ECtHR found that a death occurring at a vehicle checkpoint in south-eastern Iraq occurred within the jurisdiction of the Netherlands as it had assumed responsibility for providing security in that area and exercised its jurisdiction 'for the purpose of asserting authority and control over persons passing through the checkpoint'.[123]

In light of the above, would the extraterritorial use of force by means of an AWS amount to sufficient control of an area and/or over individuals within the AWS' sensor or weapons range to establish a jurisdictional link? It is noteworthy that in *Jaloud*, the ECtHR considered that the Netherlands exercised jurisdiction although the checkpoint was not manned by Dutch soldiers.[124] The direct involvement of agents of the state is, thus, not a necessary condition. According to Milanovic, the reasoning adopted in *Al-Skeini* would, however, exclude drone operations from the purview of human rights treaties on the basis that states using armed drones abroad do not exercise the required control over the area, nor physical control over individuals within that area, akin to taking them into custody.[125] From this standpoint, the use of an AWS would not establish the required jurisdictional link either, irrespective of whether targets are selected and attacked with or without human intervention. Melzer, on the other hand, argues that 'a State exercising sufficient factual control or power to carry out a targeted killing will also exercise sufficient factual control to assume legal responsibility for its failure to "respect" the right to life of the targeted person'.[126] Importantly, though, Melzer's statement is limited to the use of lethal force 'with the intent, premeditation and deliberation to kill individually selected persons who are not in the physical custody of those targeting them'.[127] Most other definitions of targeted killings also require intent to target (a) *specific* individual(s).[128] Whether targeting by means of an AWS can be equated with 'targeted killing' is questionable. To the extent that 'identification' of targets is predicated on an algorithmic analysis of patterns rather than the recognition of nominal identities, it is not individual but generic.[129] The extraterritorial use of force by means of an AWS may, thus, not (in all cases) establish the required jurisdictional link.

Rosén, considering drone attacks more generally, contends, in contrast, that the 'surveillance and control capabilities of drone technology … suggest a capability for exercising a degree of control and authority over territories and persons that may trigger the extraterritorial application of the [ECHR]'.[130] From this perspective, it is not the deliberate selection of individuals as targets that brings them within the jurisdiction of the state, but the 'proximity and visibility' enabled by drones, which involves a strong aspect of control. For Rosén, drones are a 'medium of proximity' and it is the capability of 'seeing and knowing' that may trigger obligations. Whether the same can be said of AWS is doubtful, however, considering that human intervention is purposefully removed from the target selection process and that human agents may therefore neither see nor know specific targets selected by the system.

On the other hand, in Rosén's account it is not the direct human intervention in the target selection process, but the *capabilities* of control offered by persistent surveillance combined with instant weapons delivery that transform the concept of

119 See England and Wales Court of Appeal (Civil Division), *Al-Saadoon & Ors v The Secretary of State for Defence & Ors* [2016] EWCA Civ 811 (09 September 2016), http://www.bailii.org/ew/cases/EWCA/Civ/2016/811.html (grappling with the question of whether the ECHR applies whenever and wherever a contracting party uses physical force). Hart points out a principled problem with requiring an element of control *in addition* to the use of force itself: such a position would imply 'that a sniper picking off a civilian at 1km would be non-justiciable, whereas soldiers cornering a group of civilians up an alley-way before swiftly despatching them would arguably give rise to a justiciable killing' (Hart, 'War Remains Inside the Court Room', supra fn 116).

120 ECtHR, *Banković et al v Belgium et al*, App no 52207/99, Grand Chamber, Decision on Admissibility, 12 December 2001, §71. The implicated governments denied having control over the airspace, and rejected that any such control could be equated with territorial control of a nature and extent that results in the exercise of effective control or of legal authority (Ibid, §44). Similarly, ECtHR, *Issa et al v Turkey*, App no 31821/96, Judgment, 16 November 2004, §82 (no jurisdiction due to failure to prove that Turkish cross-border operations amounted to the exercise of effective control of an area in Northern Iraq).

121 ECtHR, *Pad et al v Turkey*, App no 60167/00, Decision on Admissibility, 28 June 2007, §§54-55.

122 *Al-Skeini*, supra fn 100, §149.

123 ECtHR, *Jaloud v The Netherlands*, App no 47708/08, Grand Chamber, Judgment, 20 November 2014, §152.

124 According to the Dutch Government, '[a]lthough Netherlands military personnel had been there at the relevant time to observe and advise, this did not imply a hierarchical relationship such as would render the Netherlands responsible: authority rested with the Iraqi security forces' (*Jaloud*, supra fn 123, §117).

125 Milanovic, 'Al-Skeini and Al-Jedda in Strasbourg', supra fn 116.

126 Melzer, *Targeted Killing in International Law*, supra fn 92, p 139 (noting that 'the extent to which a State also has a positive obligation to actively "protect" the right to life of individuals outside its territorial jurisdiction … must be determined by reference to the level of control actually exercised over the territory or person in question').

127 Melzer, *Targeted Killing in International Law*, supra fn 92, p 5.

128 See ibid for an overview.

129 Chamayou, *Drone Theory*, supra fn 1, p 42.

130 F. Rosén, 'Extremely Stealthy and Incredibly Close: Drones, Control and Legal Responsibility', 19 *Journal of Conflict & Security Law* 1 (2014) 114, 117 (asking (at 121): 'If long surveillance periods (that today mostly lie ahead of targeted or signature killings) combined with enforcement capability in the form of instant weapons delivery does not imply an intense form of "effective control", then what constitutes "effective control"?')

responsibility under IHRL and trigger extraterritorial human rights obligations.[131] Similarly, Lieblich and Benvenisti observe that the concept of control has been significantly broadened to ensure those who exercise power bear responsibility, even if the results affect those found beyond borders where control over territory is not complete.[132] They describe 'the process of targeting' as 'a form of control *par excellence*', even if not always accepted as such in international jurisprudence.[133]

This orientation is supported by arguments in favour of a reconceptualization of human rights obligations in the digital age. Confronted with massive privacy infringements committed with secret mass surveillance programs, some suggest that the mere surveillance of individuals abroad amounts to 'virtual control' sufficient to trigger the extraterritorial applicability of human rights treaties.[134] *A fortiori*, exposing individuals to surveillance coupled with the threat or perceived risk of being made the target of attack by a machine functioning according to parameters that are unknown to those within its sensor and weapons range, and which, at any rate, they cannot influence and may not be able to escape, must bring these people within the jurisdiction of the state using the AWS.[135] Asserting extraterritorial jurisdiction is also congruent with the broader consideration that states must not be allowed to evade their responsibilities under human rights treaties by introducing a new weapon technology that reduces human control over specific force applications.

B. *ANIMUS BELLIGERENDI*: AWS AND THE INTENT TO WAGE WAR

One concern that is sometimes voiced about AWS is that they could 'accidentally trigger a war'.[136] Consider a sentry-AWS deployed in times of peace to secure an international border whose targeting parameters or sensor and weapons ranges are insufficiently restricted to prevent it from firing at foreign soldiers. Given the removal of human agents from specific force applications, could such an AWS trigger an IAC 'on its own' and thereby bring IHL into application? The question also arises in relation to non-international armed conflicts (NIACs), a topic explored by Radin and Coats in a recent article.[137] The following discussion concentrates on acts capable of triggering an IAC.

The majority view among legal scholars is that any unconsented-to military operation of a state's armed forces on the territory of another state 'could constitute a unilateral and hostile use of armed force meeting the conditions' of an IAC.[138] Most scholars agree that hostilities do not have to be of a specified level of intensity for IHL of IAC to apply.[139] In a recent in-depth study, Carron underlines that 'hostilities' capable of triggering an IAC must entail recourse to 'armed force' between states (even if there is no armed resistance).[140] She confirms that there is no requirement regarding the duration or the repetition of acts of violence, but maintains that to trigger an IAC such acts need to result in violent effects, that is, they have to cause physical harm in the form of death, injury or material damage, or serious disruption of critical infrastructures. The surveillance of military forces by another state, incursions into another state's territory or airspace, or border incidents not entailing the use of armed force would, thus, not be sufficient.[141]

That acts carried out by means of an AWS can amount to use of 'armed force' and cause physical harm is clear. What is less certain is whether the causal and intent-related requirements would be met in light of the reduced human involve-

131 Ibid, 122.

132 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81, p 263–264.

133 Ibid, p 264.

134 A. Peters, 'Surveillance without Borders: The Unlawfulness of the NSA Panopticon, Part II', *EJIL: Talk!*, 4 November 2013, http://www.ejiltalk.org/surveillance-without-borders-the-unlawfulness-of-the-nsa-panopticon-part-ii/ (arguing that '[i]t is not too far-fetched in the cyber-age to imagine that this type of control ['virtual control' due to mere surveillance] might also trigger the human rights obligations of the "virtual" controller'). See also C. Nyst, 'Interference-Based Jurisdiction Over Violations of the Right to Privacy', *EJIL: Talk!*, 21 November 2013, http://www.ejiltalk.org/interference-based-jurisdiction-over-violations-of-the-right-to-privacy/.

135 On the significant adverse impacts of armed drones on peoples' lives, physical and mental health and livelihoods, see International Human Rights and Conflict Resolution Clinic (Stanford Law School) and Global Justice Clinic (NYU School of Law), *Living Under Drones: Death, Injury, and Trauma to Civilians From US Drone Practices in Pakistan*, September 2012, http://chrgj.org/wp-content/uploads/2012/10/Living-Under-Drones.pdf.

136 E.g., A. Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons*, Ashgate, 2009, 152.

137 S. Radin and J. Coats, 'Autonomous Weapon Systems and the Threshold of Non-International Armed Conflict', 30 *Temple International and Comparative Law Journal* (2016) 133–150, available at SSRN, https://ssrn.com/abstract=2887130.

138 ICRC, Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (ICRC Commentary GC I), 2nd edn, 2016, §241, https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=BE2D518CF5DE54EAC1257F7D0036B518.

139 In contrast, Committee on the Use of Force, *Final Report on the Meaning of Armed Conflict in International Law (Summary)*, International Law Association, 2010, p 2 (stating that '[t]he violence must be organized and intense – even between sovereign states – before the otherwise prevailing peacetime rules are suspended'). With respect to the threshold for the application of IHL of NIAC, the majority view is that 'the violence needs to have reached a certain intensity and that it must be between at least two organized Parties/armed groups' (ICRC Commentary GC I (2016), supra fn 138, §421). For a different view, see A. A. Haque, 'Triggers and Thresholds of Non-International Armed Conflict', *Just Security*, 29 September 2016, https://www.justsecurity.org/33222/triggers-thresholds-non-international-armed-conflict (arguing that 'if an armed group is sufficiently organized, then a first use of armed force by or against that group should trigger a NIAC'). At what stage a state using an AWS finds itself in a situation of occupation governed by IHL of IAC is a question beyond the scope of this study. Note that control exercised by a state sufficient to bring persons abroad within its jurisdiction, possibly, by means of an AWS (see above) does not necessarily reach the threshold of an occupation (see Droege, 'The Interplay', supra fn 108, 332).

140 D. Carron, 'L'acte déclencheur d'un conflit armé international', Thèse de doctorat no D. 902, Université de Genève, 2015, pp 218, 304, http://archive-ouverte.unige.ch/unige:75120.

141 Ibid, pp 210, 212, 218, 238. See also Roscini, *Cyber Operations*, supra fn 41, p 136. Carron thereby rejects another prevalent interpretation according to which IHL of IAC applies as soon as events or persons exist that fall within the purview of the Geneva Conventions and AP I (e.g. soldiers captured by another state's soldiers). She finds ('L'acte déclencheur d'un conflit armé international', supra fn 140, p 217) that it is hostilities between or among states that create the IAC and that it is only once an IAC exists that people become 'protected persons' under the Geneva Conventions. According to Carron, use of force intended to cause physical effects, but which fails to do so, for whatever reason, does not trigger an IAC. However, such acts of violence can amount to an 'attack' in the sense of Art 49, AP I, once an IAC exists. On the interpretation of the IHL notion of 'attack', see the next section.

ment in the use of force.[142] For an IAC to be triggered, the use of force has to be carried out by state agents or other persons authorized to act on the state's behalf, and there has to be a direct causal link to the state's 'intent' to engage in hostilities against another state. This so-called *animus belligerendi* tends to be expressed in instructions to state agents. Although debate persists on this point, from this perspective, which is shared by the ICRC, situations that are the result of a mistake or of *ultra vires* acts[143] do not trigger an IAC.[144]

Establishing the *animus belligerendi* is particularly challenging when weapons are involved whose violent effects are spatially distributed or temporally deferred, for instance, because they are 'victim-activated', such as munitions designed to be exploded by the presence, proximity or contact of a person or vehicle: mines.[145] Carron considers that the laying of mines does not in itself trigger an IAC because violent effects are not produced by that act. She acknowledges, however, that the detonation of a mine at a later point in time may not trigger an IAC either because the causal link to the state's intent to attack another state may be too remote.[146]

Many commentators have pointed to challenges in attributing the acts of an AWS to a state (a prerequisite for holding it responsible for violations of international law)[147] precisely because its applications of force can be spatially, temporally and causally remote from a state agent's decision to use force.[148] If we accept that there needs to be a proximate causal link between an act of violence and a state's intent to conduct hostilities against another state, the lack of human control exercised over *specific* applications of force in the use of an AWS means that an AWS that is *not* specifically deployed to engage in hostilities cannot trigger an IAC 'on its own' (even if a state would *a priori* be responsible for all acts of an AWS carried out *in the course of an armed conflict*). A sentry-AWS installed for border control purposes that fires at members of a neighboring state's border guard, would, thus, not trigger an IAC. Consequently, such acts remain governed by IHRL standards on the use of force.[149]

The conclusion that an AWS cannot 'accidentally trigger a war' is congruent with the broader consideration that, as an adjunct of the right to life, any doubt about the existence of an armed conflict is to be resolved in favour of peace.[150] At the same time, however, Carron points out that the state whose people or objects come under attack from another state may justifiably presume that the use of armed force is intentionally hostile.[151] Even if an AWS cannot trigger an IAC 'on its own', its deployment bears a real risk of escalation.

C. THE BELLIGERENT NEXUS: AWS, CONTROL AND THE INTENT TO CONDUCT HOSTILITIES

Like certain other technologies and practices of violence, such as hostile activities in the cyber domain or the use of armed drones for 'targeted killings' abroad in the fight against 'terrorism' or 'violent extremism', AWS challenge traditional notions around which international legal standards on the use of force are articulated. The increasing expectation on armed forces that they not only conduct combat operations but also fulfill law enforcement tasks has called into question the distinction between war-fighting and policing. Certain activities, such as enforcing a roadblock, can in one instance be part of hostilities and in another be part of law enforcement. In addition, there is growing convergence of military and policing technologies. This raises the question of whether (that is, in what circumstances) the use of a sentry-AWS during an armed conflict to secure a perimeter around a detention camp, a checkpoint or a military base, would be governed by the normative paradigm of hostilities, rather than being assessed within a law enforcement framework. As we shall see, the exercise of human control is required in order to use an AWS as a means of warfare but, at the same time, control over the context within which the use of force takes place limits the application of the conduct of hostilities paradigm.

'Hostilities' are sometimes described as the (collective) resort by parties to an armed conflict to means and methods of injuring the enemy.[152] They comprise 'all activities, which are designed to support one party to the conflict by harming an-

142 Another relevant aspect for the determination of whether an act of violence triggers an IAC relates to who or what the violence is directed at (the target). Although not discussed here, this raises questions about the target detection and selection parameters of the AWS.

143 ICRC Commentary GC I (2016), supra fn 138, §241.

144 Carron, 'L'acte déclencheur d'un conflit armé international', supra fn 140, pp 353, 369.

145 Art 2(1), CCW AmPII, supra fn 86. An anti-personnel mine is defined as 'a mine designed to be exploded by the presence, proximity or contact of a person and that will incapacitate, injure or kill one or more persons' in Art 2(1), APMBT, supra fn 88.

146 Carron, 'L'acte déclencheur d'un conflit armé international', supra fn 140, fn 1329, p 239. In contrast, a civilian who lays a mine in the course of an armed conflict would be considered to directly participate in 'hostilities' (DPH Guidance, supra fn 106, p 55). At what point an attack in the sense of Art 49 AP I exists is addressed in the following section. For a discussion about an 'armed attack' under *jus ad bellum* in connection with automated weapon systems, see Grimal, 'Missile Defence Shields', supra fn 89, 5.

147 For an introduction, see T. Marauhn, 'An Analysis of the Potential Impact of Lethal Autonomous Weapons Systems on Responsibility and Accountability for Violations of International Law', Presentation, CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 13-16 May 2014, https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/assets/media/35FEA015C2466A57C1257CE4004BCA51/file/Marauhn_MX_Laws_SpeakingNotes_2014.pdf. For a more detailed analysis, see T. Chengeta, *Accountability Gap, Autonomous Weapon Systems and Modes of Responsibility in International Law*, 30 September 2015, available at SSRN, http://dx.doi.org/10.2139/ssrn.2755211.

148 Although the analogy is not perfect (as AWS could behave less deterministically and could search for targets more actively than mines), sentry-AWS, especially stationary ones, can be likened to mines in that 'the human agent directly identifiable as the efficient cause of death' is the victim (Chamayou, *Drone Theory*, supra fn 1, p 211).

149 As noted earlier, such acts can violate *jus ad bellum*. It is also worth recalling that under IHRL, a government is not free to escalate its use of force in order to create a NIAC (A. Bellal and L. Doswald-Beck, 'Evaluating the Use of Force During the Arab Spring', in M. Schmitt and L. Arimatsu (eds), 14 *Yearbook of International Humanitarian Law* (2011) 32).

150 M. E. O'Connell, 'Remote-Controlled Killing in Dallas', *EJIL: Talk!*, 19 July 2016, http://www.ejiltalk.org/remote-controlled-killing-in-dallas/.

151 Carron, 'L'acte déclencheur d'un conflit armé international', supra fn 140, p 364.

152 DPH Guidance, supra fn 106, p 43. The notion of 'hostilities' or related notions like 'military operations', 'combat' or 'warfare' are not expressly defined under IHL.

other'.[153] The 'law of hostilities' accordingly consists of those rules and principles that govern the 'choice of and use by the parties to an armed conflict of means and methods of injuring the enemy.'[154] So, when does the use of an AWS constitute use of a 'means of injuring the enemy', a 'means of combat' or a 'means of warfare'?[155]

'Means of warfare' have been described in the context of air and missile warfare, as 'weapons, weapon systems or platforms employed for the purposes of attack'.[156] This includes objects upon which an attacking platform directly relies to carry out an attack such as system components that 'provide targeting data and other essential information' to a platform actually engaging a target. In contrast, components that contribute to military operations, but are 'not designed or used to injure, kill or damage enemy personnel or objects' are not included.[157] Similarly, the *Tallinn Manual* describes 'cyber means of warfare' as including any 'cyber device, materiel, instrument, mechanism, equipment, or software used, designed or intended to be used to conduct a cyber attack'.[158] Insights from discussions on cyber security are of special interest to the debate on AWS as cyber and autonomous weapons intersect and raise some common challenges to the normative regulation of and human control over the use of force.[159]

Implicit in this conception of a 'means of warfare' is, first, the requirement of a 'belligerent nexus'. In other words, for a weapon system to be governed by the law of hostilities its use has to be 'designed to support one party to an armed conflict against another'.[160] Establishing that (design-)intent may be challenging when force is used by means of an AWS because the belligerent nexus is context-dependent[161] and specific applications of force may not be under human control. Second, the concept of a 'means of warfare' is tied to the IHL notion of 'attack'.

153 Melzer, *Targeted Killing in International Law*, supra fn 92, p 276.

154 Ibid, p 269.

155 The term 'means of warfare' is used, e.g., in Art 36, AP I. Art 51(4), AP I refers to 'means of combat'. Art 22, 1907 Hague Regulations refers to 'means of injuring the enemy'.

156 HPCR, AMW Manual, supra fn 42, Rule 1(t).

157 HPCR, Commentary on AMW Manual, supra fn 42, Rule 1(t), §4, p 42.

158 Schmitt, *Tallinn Manual*, supra fn 42, Commentary on Rule 41, §2, p 142.

159 See, e.g., 'Cyber Weapons and Autonomous Weapons: Potential Overlap, Interaction and Vulnerabilities', Conference, UNIDIR, 9 October 2015, http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-addressing-competing-narratives-phase-ii/cyber-weapons-and-autonomous-weapons-potential-overlap-interaction-and-vulnerabilities.

160 Melzer, *Targeted Killing in International Law*, supra fn 92, p 276; Roscini, *Cyber Operations*, supra fn 41, pp 123–124; DPH Guidance, supra fn 106, p 58.

161 In determining whether an act is sufficiently related to an armed conflict so as to amount to a possible violation of IHL, the International Criminal Tribunal for the Former Yugoslavia (ICTY) has had regard to 'the fact that the perpetrator is a combatant; … the victim is a non-combatant; … the victim is a member of the opposing party; … the act may be said to serve the ultimate goal of a military campaign; and … that the crime is committed as part of or in the context of the perpetrator's official duties' (ICTY, *The Prosecutor v Kunarac, Kovač and Vuković*, IT-96-23&IT-96-23/1-A, Appeals Chamber, Judgment, 12 June 2002, §59). However, none of these indicators is on its own conclusive (Roscini, *Cyber Operations*, supra fn 41, p 125; Gaggioli and Kolb, 'A Right to Life in Armed Conflict?', supra fn 109, 47) .

Attacks are defined under IHL as 'acts of violence against the adversary, whether in offence or in defence'.[162] Recent commentaries clarify that this includes 'operations that actually result in violent effects, and those which were intended to but failed'.[163] Acts of violence directed at civilians (unlawfully) are also attacks.[164] Although the precise temporal and geographic relation of military operations, hostilities, attacks and targets is not well established,[165] most commentators consider that an attack 'as a whole' can encompass a series of incidents or engagements.[166] Whereas the relationship between 'hostilities' and 'military operations' remains somewhat ambiguous, both notions are said to include 'attacks'. In addition to attacks, military operations can also include other activities directly connected to the use of a weapon or weapon platform involving the actual or potential use of force against an enemy, as well as operations in direct support of such operations.[167] Under IHL, civilians and civilian objects are protected from unlawful attacks, and they 'enjoy general protection against dangers arising from *military operations*'.[168] However, according to a significant part of legal scholarship, the rules on targeting (proportionality, distinction and precautions in attack) apply to attacks only.[169] So, what constitutes 'an attack' in the use of an AWS?

162 Art 49, AP I.

163 HPCR, Commentary on AMW Manual, supra fn 42, Rule 1(e)(1) and (6); Schmitt, *Tallinn Manual*, supra fn 42, Commentary on Rule 30, §§7 and 15, 110.

164 E.g. International Criminal Court (ICC), *The Prosecutor v Bosco Ntaganda*, ICC-01/04-02/06, Decision Pursuant to Article 61(7)(a) and (b) of the Rome Statute (Pre-Trial Chamber II), 9 June 2014, §§45–48.

165 R. C. Else, 'Proportionality in the Law of Armed Conflict: The Proper Unit of Analysis for Military Operations', Note, 5 *University of St. Thomas Journal of Law & Public Policy* 1 (2010) 208.

166 See the reservation to this effect made by several states upon ratification of AP I, e.g. the reservations by the UK to Arts 51 and 57 AP I, 28 January 1998, https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Notification.xsp?action=openDocument&documentId=0A9E03F0F2EE757CC1256402003FB6D2. See also A. Jachec-Neale, *The Concept of Military Objectives in International Law and Targeting Practice*, Routledge, 2015, pp 121–122; ICTY, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia*, §78, http://www.icty.org/en/press/final-report-prosecutor-committee-established-review-nato-bombing-campaign-against-federal.

167 According to the HPCR Commentary on the AMW Manual (supra fn 42, Rule 1(b)(1)–(c)(3), pp 25–27) 'military operations' include attacks, interceptions, as well as 'activities directly connected to the actual use of the aircraft or missile such as deployment, launching, guidance or retrieval', and involve 'actual or potential use of force against an enemy; and (ii) operations in direct support of the aforementioned operations'. 'Air or missile *combat* operations' (emphasis added) mean 'air or missile operations designed to injure, kill, destroy, damage, capture or neutralize targets, the support of such operations, or active defence against them'. They include attacks as well as 'refueling; jamming of enemy radars; suppression of enemy defences by attacking enemy radar stations and anti-aircraft artillery or missile sites; use of airborne warning and control systems; bombing; fighter escort and fighter sweeps preceding bomber attacks' (Rule 1(c)(3)).

168 Arts 51(1) and 57(2), AP I (emphasis added)

169 Consider, for example, the obligation on parties to conflict to exercise 'Control during the Execution of Attacks' (ICRC CIHL Database, Rule 19, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule19), so as to take all feasible measures to cancel or suspend 'an attack' if it becomes apparent that it is not directed at a legal target or may be expected to cause disproportionate civilian harm. Some scholars consider, in contrast, that the rules on targeting apply not to 'attacks', but rather to the broader notion of 'hostilities'. For a brief overview of this controversy, see Roscini, *Cyber Operations*, supra fn 41, p 181.

As mentioned earlier, the violent effects of an AWS, not unlike those of mines, can be deferred in space and time. When such weapons are involved in the use of force, when does an attack begin and end? One approach, reportedly adopted by the drafters of Article 49 of AP I in relation to mines, is to consider that 'there is an attack whenever a person is directly endangered by a mine laid'.[170] At what point a person is 'directly endangered' by a mine was not clarified, though. Depending on how 'directly endangered' is interpreted, 'an attack' can be understood quite narrowly. Applied to AWS, that could mean that every instance in which an AWS selects (or even only detects) a target, the person or object concerned is 'directly endangered' by the AWS. Consequently, every such instance would individually constitute 'an attack' to which the rules on targeting apply. In light of IHL restrictions and requirements pertaining to attacks, such an approach would impose a degree of human control over individual applications of force that leaves little to no room for the use of AWS.[171]

Another approach is to consider that 'an attack' starts with the activation of an AWS to combat another party to an armed conflict,[172] and to treat all persons and objects that potentially fall within its target parameters as being 'directly endangered'. In this case, the belligerent nexus can be presumed for subsequent acts of violence committed with the AWS, but this wider notion of 'attack' raises significant concerns about compliance with IHL rules on targeting: if the AWS has broad targeting parameters and operates independently over a wide area and a long timespan, how can compliance with targeting rules be ensured, even though the number and context of *specific* acts of violence may not be known when 'the attack' is launched?

The uncertain spatio-temporal boundaries of 'an attack' and a tendency among some legal commentators to 'shift back the point of assessment to the decision to deploy the weapon',[173] give rise to the concern that developments in weapon technologies result in a continuous expansion of the concept of attack (and of hostilities more generally). If the law is to function, however, there 'has to be some spatial, temporal, or conceptual boundaries to an attack',[174] and indeed, some limitations are implicit in IHL.[175] Ongoing debates about the appropriate level at which a military objective should be defined in the cyber context,[176] and on ways to counteract a trend towards assessing proportionality in the aggregate, suggest that explicit restrictions may be called for.[177]

Conversely, if an AWS was *not* activated in order to support one party to an armed conflict against another, as may be the case with an autonomous sentry system installed to secure the perimeter around a power plant, detention facility or along an international border, the belligerent nexus cannot be presumed for subsequent applications of force. So, when is the AWS used as a means of warfare to conduct an attack? In the context of cyber operations it has been proposed that the rules on attacks apply in relation to a party to an armed conflict that 'controls' or acquires 'sufficient control' over a weapon (system) to employ it as if it were its own,[178] and that 'an object must be in the control of an attacking party to comprise a means of warfare'.[179] The Internet, for instance, is not a 'means of warfare' even if it connects an attacker's computer system to a target. Arguably, therefore, agents of a party to an armed conflict have to exercise sufficiently proximate human control over the AWS to establish the required belligerent nexus for subsequent force applications to be governed by the conduct of hostilities paradigm.

The requirement to exercise control points to a related legal debate; that about potential legal limits on where and when hostilities, including attacks, may take place. Like certain other new weapon technologies, AWS could enable the use of force in areas and over timespans that are not easily covered by human operators.[180] Although IHL does not provide for spatial limitations on where hostilities may take place, many have warned against treating the entire world as a 'global battlefield' in

170 ICRC, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (ICRC Commentary APs), Martinus Nijhoff Publishers, 1987, §1881, p 603.

171 Of course, such a restrictive interpretation may well be justified, considering that an AWS with mobile components, a vast sensor array and operating within broad parameters could potentially endanger a lot more people than a landmine.

172 In this vein, R. Sparrow, 'Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender', 91 *International Law Studies* (2015) 725: 'If AWS are weapons then launching an AWS is launching an attack. Moreover, it seems natural to think of this as launching an attack against *all* of the targets that the AWS might in fact strike' (original emphasis).

173 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81, p 255.

174 Article 36, *Key Elements of Meaningful Human Control*, Background paper to comments by Richard Moyes for the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), April 2016, p 3, http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf.

175 For example, the principle of military necessity has a restrictive dimension from which derive the requirements of 'effective contribution', 'definite military advantage' (Art 52(2), AP I), and 'concrete and direct military advantage' (Art 57(2)(a)(iii), AP I).

176 H. Harrison Dinnis, 'The Nature of Objects: Targeting Networks and the Challenge of Defining Cyber Military Objectives', in 48 *Israel Law Review* 1 (2015) 50–54 (discussing a potential requirement to define a military objective at the most specific level (in its most minimal form) in the cyber context).

177 Else, 'Proportionality in the Law of Armed Conflict', supra fn 165, 195–213 (proposing a requirement of temporal and geographic proximity to assess proportionality when the military advantage depends on damaging a series of targets). In contrast, see ICC, *Elements of Crimes*, 2011, fn 36, p 19.

178 Schmitt, *Tallinn Manual*, supra fn 42, Commentary on s 5, §2, p 141.

179 Ibid, Commentary on Rule 41, §3, p 142. It should be kept in mind that a weapon system that is used as a means of warfare will in all but exceptional situations qualify as a military objective by its nature pursuant to Art 52(2), AP I, and may, thus, be attacked. Harrison Dinniss convincingly argues that an AWS' sensor array, 'code' and cyber infrastructure making up its network and databases, irrespective of their intangibility, are objects in the sense of Art 52(2), AP I. Where an AWS makes use of civilian networks (servers, fibre-optic cables, etc.) this exposes 'vast amounts of that infrastructure to attack' and raises questions about 'the precise level at which the military objective should be defined – code, component, system or network level' (Harrison Dinnis, 'The Nature of Objects', supra fn 176, 46–48, 50). This question is connected to that about the spatio-temporal boundaries of an 'attack' (see below).

180 UNIDIR, *The Weaponization of Increasingly Autonomous Technologies in the Maritime Environment: Testing the Waters*, pp 3, 5, http://www.unidir.ch/files/publications/pdfs/testing-the-waters-en-634.pdf (noting that '[a]utonomous technologies will make possible "lay and wait" (so-called "long-loiter") missions of hitherto unimagined duration').

an 'everywhere' and 'forever war',[181] and some suggest that AWS should only ever be operated in 'regions of heavy fighting', 'kill boxes' or 'engagement regions'.[182] Would it be permissible, for example, for a party to an armed conflict to direct force against a person by means of a sentry-AWS in a location far from the 'heart of the battlefield' even if that person would be a legitimate target under IHL?[183]

It is worth recalling that the law enforcement model is the default paradigm. Whether it can reasonably be applied depends on the context within which violent effects are produced. Gaggioli and Kolb propose that the conduct of hostilities model is applicable as *lex specialis* if persons are targeted who are legitimate targets of attack under IHL, the State is deprived of sufficient control over persons to enable arrest and the degree of violence involved is high.[184] Whereas some consider that 'the status, function or conduct of the person against whom force may be used' is the main criterion to decide whether a use of force is to be assessed within a law enforcement or a conduct of hostilities paradigm,[185] others, including the ICRC in its *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law*, take into consideration the ability of a party to the conflict 'to control the circumstances and area' where force is used.[186] From the latter standpoint, it can be argued that the more control a party exercises over a situation, the more the military necessity to use force under a conduct of hostilities paradigm diminishes.[187] Gaggioli and Kolb underline in this respect that '[t]he control at stake is ... a factual control over the individual, determining if it is materially feasible to proceed to an arrest', even on territory not controlled by the belligerent.[188]

181 See, e.g., D. Gregory, 'The Everywhere War', 177 *The Geographical Journal* 3 (2011) 238–250, doi: 10.1111/j.1475-4959.2011.00426.x. On the 'spatial dynamics of post-modern warfare', see Bolton, 'From Minefields to Minespace', supra fn 87, 43–44, with further references.

182 Lin et al, *Autonomous Military Robotics*, supra fn 72, p 77.

183 For a discussion in the context of armed drones, see N. Lubell and N. Derejko, 'A Global Battlefield? Drones and the Geographical Scope of Armed Conflict', 5 *Journal of International Criminal Justice* 1 (2013) 65–88, doi: 10.1093/jicj/mqs096; J. Pejic, 'Extraterritorial Targeting by Means of Armed Drones: Some Legal Implications', 66 *IRRC* 893 (2014) 67–106, https://www.icrc.org/en/download/file/7375/jelena_pejic-_armed_drones_-_final_pdf.pdf.

184 Gaggioli and Kolb, 'A Right to Life in Armed Conflict?', supra fn 109, 47.

185 For many experts participating in a meeting organized by the ICRC on this topic in 2012, 'the main (if not the only) legal criterion for determining what paradigm governs the use of force is the status, function or conduct of the person against whom force may be used' (Gaggioli, *The Use of Force in Armed Conflicts*, supra fn 111, p 59). From this standpoint, it would be legal to direct an attack against a person with combatant status under IHL, irrespective of that person's location.

186 DPH Guidance, supra fn 106, p 80. Similarly, Bellal and Doswald-Beck, 'Evaluating the Use of Force During the Arab Spring', supra fn 149, 14 (noting that 'the evaluation of the lawfulness of any particular use of force will depend on the degree of control over territory or over a person in both human rights law and humanitarian law').

187 See also Melzer, *Targeted Killing in International Law*, supra fn 92, p 297. Control over an area, the intensity of violence or whether the situation is taking place inside or outside a 'conflict zone', 'battlefield' or 'zone of operations' were, however, not seen as decisive by most experts participating in the 2012 meeting (Gaggioli, *The Use of Force in Armed* Conflicts, supra fn 111, p 59).

188 Gaggioli and Kolb, 'A Right to Life in Armed Conflict?', supra fn 109, 47.

Although scholars may not agree on the legal basis for potential spatial limitations on the conduct of hostilities, as a practical matter, many advocate an escalation of force procedure in contentious situations,[189] where force is used depending on the threat posed by the target, rather than its status or function.[190] In light of the fluidity of contemporary armed violence, human agents involved in the use of an AWS need to be in a position to recognize when control over circumstances, an area or an individual enable and thus require the application of law enforcement standards, and to adapt operations accordingly.

D. PRELIMINARY FINDINGS ON THE APPLICABLE LAW

The reconfiguration of the human–machine relationship that accompanies increasing automation in weapon systems raises concerns about the ability of human agents involved in the use of force by means of an AWS to comply with legal rules for the protection of the human person. Increasing 'autonomy in critical functions' comes at the price of reduced predictability in the use of force and challenges to ensuring accountability for its consequences. The mode of human intervention in the use of force can also affect what legal rules apply. The spatial, temporal and causal remoteness of human intervention in the use of an AWS from the locus of force application impacts intent- and control-related determinants of the applicable law. From the standpoint that AWS do not themselves make legal determinations, compliance with the law demands, among other things, that in the use of an AWS, human agents exercise the control necessary to determine what legal rules govern the use of force in specific circumstances.

Through the prism of three ongoing legal debates, the discussion above has brought out the tensions that can exist between the expectation on states – more specifically their human agents – that they control the use of weapons, the acceptance that this expectation cannot extend to matters beyond the state's factual control, and the assertion of legal control in situations where evolving practices of violence risk undermining the object and purpose of legal rules for the protection of the human person. In such situations, the assertion of legal control can be a strong incentive for states to assume factual control. Specifically, whether the use of force abroad by means of an AWS amounts to control over an area or individual sufficient to establish a jurisdictional link for the extraterritorial application of IHRL treaties is uncertain, especially if one adopts the position that the human involvement in the violence is too remote and/or that there is no intent to target specific individuals. On the other hand, a compelling argument can be made that the persistent surveillance and instant weapons delivery enabled by the use of an AWS presents capabilities of control strong enough to bring those within the AWS' sensor and weapons range within the jurisdiction of the user state, and thereby within the sphere of protection of the IHRL treaties the state is bound by.

189 Gaggioli, *The Use of Force in Armed* Conflicts, supra fn 111, pp 59–60.

190 An approach endorsed by Arkin, *Governing Lethal Behavior*, supra fn 81, p 11.

Lack of proximate human involvement must also mean that an AWS cannot trigger a war, 'on its own' and thereby bring IHL of IAC into operation. In the absence of an explicit expression by human state agents of the will to 'wage war' against another state, the state's *animus belligerendi* cannot be presumed. A sentry-AWS deployed to secure an international boundary in times of peace, for example, cannot 'accidentally' trigger an IAC. Its applications of force thus remain governed by IHRL standards on the use of force.

Likewise, during an armed conflict, for the law of hostilities to govern the use of force by means of an AWS, there needs to be a 'belligerent nexus', that is, the use of force must be designed or intended by human agents to serve one party to the conflict by harming another. If the AWS was not activated by a human agent with conduct of hostilities in mind, this belligerent nexus cannot be presumed and applications of force remain governed by the law enforcement paradigm. In order to use an AWS to conduct hostilities, human agents of a party to the conflict have to exercise sufficiently proximate control over the system to use it as a means of warfare. Conversely, if an AWS is activated by a human agent of a party to an armed conflict with the intent to conduct hostilities, the belligerent nexus can be presumed for subsequent applications of force. However, to ensure that targeting rules can be applied so as to provide effective protection to the victims of war, even though the number and context of specific acts of violence may not be known when an attack is launched, human agents have to bound an attack appropriately in spatio-temporal terms. Furthermore, if a state, by means of an AWS or otherwise, exercises control over the context within which violent effects are produced, including persons that force is directed at, so that a differentiated use of force following a law enforcement logic becomes possible, an AWS can no longer operate according to a conduct of hostilities model.

Despite the focus on IHL in policy discussions and commentators envisioning AWS operating in empty spaces far away,[191] IHL is not the only legal frame of reference for the use of force by means of an AWS, and may in a number of situations not be the primary one. To judge from contemporary armed violence situations, there may not be many scenarios in which it can unquestioningly be assumed that a conduct of hostilities approach is the appropriate model for using force by means of a sentry-AWS, even during an armed conflict or along a border separating states that are technically (still) 'at war'. Consequently, human involvement in the use of force by means of an AWS must be such that human agents can determine in a timely manner when and where the law of hostilities is no longer the appropriate frame of reference and can adapt operations accordingly.[192]

Finally, the discussion above draws attention to major legal uncertainties and controversies concerning questions relevant to the determination of the applicable law. The likelihood that AWS use complies with these legal standards is, thus, not solely a question of the sophistication of algorithms and sensors, but depends, critically, on reaching a widely shared agreement on what law applies in specific circumstances and what that law demands.[193] Faced with the indeterminacy of legal norms and legal controversies, there is a tendency to defer judgement and assess issues on a case-by-case basis. It should be kept in mind, however, that 'the context' cannot resolve questions about the meaning and appropriateness of certain activities because what constitutes the appropriate context is itself contingent on social agreement.[194]

191 Schmitt and Thurnher, '"Out of the Loop"', supra fn 19, 246, 250 (portraying as *a priori* unproblematic 'the employment of such systems for an attack on a tank formation in a remote area of the desert or from warships in areas of the high seas far from maritime navigation routes', whilst acknowledging that an AWS would have to be 'capable of geographic restriction' and 'temporal limitation since few areas are always completely devoid of civilians or civilian objects'. See also W. Boothby, 'Some Legal Challenges Posed by Remote Attack', 94 *IRRC* 886 (2012) 585.

192 It has been proposed that in circumstances where it is unclear what paradigm governs the use of force, militaries thinking of deploying autonomous robots 'could simply programme them with the more restrictive of the two frameworks – international human rights law – which could continue to govern their actions even if the paradigm changed', or 'that autonomous weapons are only deployed in circumstances that unequivocally fall into one or the other paradigm', or that 'the use of autonomous robotic weapons should be restricted to the [conduct of hostilities] paradigm' (A. Leveringhaus and G. Giacca, *Robo-Wars: The Regulation of Robotic Weapons*, Oxford Martin Policy Paper, 2014, p 14, http://www.oxfordmartin.ox.ac.uk/downloads/briefings/Robo-Wars.pdf). Aside from the questionable presupposition that entire bodies of law or 'paradigms' could adequately be encoded into weapon designs, it should be kept in mind that IHRL is not in all respects more protective than IHL, and that in situations of parallel application, IHL and IHRL interact in various, complex ways. (For a nuanced discussion, see Gaggioli and Kolb, 'A Right to Life in Armed Conflict?', supra fn 109.) In the view of this author, the proposal to limit the deployment of AWS to situations in which one or the other paradigm applies, or to restrict deployment to the conduct of hostilities, entails that human agents must exercise control over the use of force necessary to become aware of changes that affect the legal qualification of a situation and make the appropriate adjustments.

193 'If there is no agreement on what constitutes an armed conflict, no agreement on who counts as a combatant, and no agreement on what constitutes an imminent threat, the law is no longer a guidepost' (R. Brooks, 'Drones and the International Rule of Law', 28 *Ethics & International Affairs* 1 (2014) 98).

194 Rappert, *Controlling the Weapons of War*, supra fn 40, pp 91–92, 102.

5. HUMAN RIGHTS REQUIREMENTS AND CONSTRAINTS ON THE USE OF AWS

In a joint report to the Human Rights Council in February 2016, the Special Rapporteur on the rights to freedom of peaceful assembly and of association, Maina Kiai, and the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, recommended that '[a]utonomous weapons systems that require no meaningful human control should be prohibited'.[195]

The Special Rapporteur on extrajudicial killings had already questioned, in a 2013 report to the Human Rights Council, whether AWS use 'is in principle acceptable, because it entails non-human entities making the determination to use lethal force'. This question, he argued, 'is an overriding consideration: if the answer is negative, no other consideration can justify the deployment of [AWS], no matter the level of technical competence at which they operate.'[196] Heyns further argued that 'the same principled issues' arise independent of 'whether the force used [by means of an AWS] is lethal or not, and whether it is used in war or policing'.[197] The remainder of this study investigates some of these 'principled issues' further and examines the extent to which they are extenuated in situations governed by IHL, including the conduct of hostilities.

A. AUTOMATED KILL ZONES: PREPARING THE GROUND FOR SENTRY-AWS?

To control their international border and regulate the movement of people and goods within them is a sovereign prerogative of states. Security and safety considerations may justify and even require that the authorities limit the public's access to certain locations, such as ammunition depots or nuclear power plants, and that they institute control measures at internal administrative boundaries and at international borders.[198] At the same time, measures taken by state agents to guard or defend a perimeter in order to detect, intercept, identify, and in some circumstances, detain or remove a person from an area can adversely affect the enjoyment of a range of human rights and the dignity of persons attempting to cross the boundary or who find themselves in or near the area for other reasons. International borders in particular, and even more so, ceasefire lines, can be dangerous places.[199] In some cases, national regulations characterize boundary areas as zones of exclusion or exception regarding human rights protections, legitimizing practices that effectively create 'zones of lawlessness' in violation of international law.[200]

A border control regime that was found to violate international law by several international human rights bodies was that instituted by the authorities of the German Democratic Republic (GDR) between 1961 and 1989. Many people trying flee to the Federal Republic of Germany lost their lives attempting to cross that border by triggering an anti-personnel mine or an 'automatic-fire system', or after being shot by East German border guards.[201] Using arguments that bear a striking resemblance to those advanced in favour of autonomous sentry-systems, the former head of the GDR's Border Troops pointed out in proceedings before the United Nations Human Rights Committee (HRCttee) that, at the time, international law did not prohibit the installation of mines along an international border. He underlined that the mines were only used 'in military exclusion zones', were 'clearly indicated by warning signs', 'involuntary access was prevented by high fences', and the danger of entering the area was known to people attempting to cross the border.[202] He argued that the mines were 'a preventive military measure against a

195 UN doc A/HRC/31/66, supra fn 16, §67(f).

196 UN doc A/HRC/23/47, supra fn 13, §§92–93.

197 Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 355.

198 'A boundary is essentially a line of definition, while a border is usually a more complex entity comprising several lines and/or zones, whose primary function is the regulation of movement of people and goods' (OSCE, *Applied Issues in International Land Boundary Delimitation/Demarcation Practices*, 2011, p 8, http://www.osce.org/cpc/85263?download=true). States can exercise border governance measures along the politically defined boundaries separating their territory or maritime zones from that of other political entities, as well as at checkpoints and border posts at train stations, sea- or airports, in transit zones and embassies, and in so-called 'no-man's land' between border posts, on their territory and extraterritorially (Office of the UN High Commissioner for Human Rights (OHCHR), *Recommended Principles and Guidelines on Human Rights at International Borders*, [undated], p 4, http://www.ohchr.org/Documents/Issues/Migration/OHCHR_Recommended_Principles_Guidelines.pdf).

199 IACommHR, 'IACHR Condemns the Recent Death of Mexican National by U.S. Border Patrol Agents', Press Release no 93/12, 24 July 2012, https://www.oas.org/en/iachr/media_center/PReleases/2012/093.asp; HRW, 'US/Mexico: Investigate Border Killings', 11 June 2010, https://www.hrw.org/news/2010/06/11/us/mexico-investigate-border-killings; B. Adams, 'India's shoot-to-kill policy on the Bangladesh border', *The Guardian*, 23 January 2011, https://www.theguardian.com/commentisfree/libertycentral/2011/jan/23/india-bangladesh-border-shoot-to-kill-policy.

200 OHCHR, *Recommended Principles*, supra fn 198, pp ii, 2.

201 ECtHR, *Streletz, Kessler and Krenz v Germany*, App nos 34044/96, 35532/97 and 44801/98, Grand Chamber, Judgment, 22 March 2001, §13. References to '*Selbstschussanlagen*' or '*Todesautomaten*' in official documents contributed to the public perception that the border was secured by automated sentry guns. In reality, these were fence-mounted, directional SM-70 fragmentation mines, triggered by trip-wire.

202 HRCttee, *Klaus Dieter Baumgarten v Germany*, Comm no 960/2000, UN doc CCPR/C/78/D/960/2000 (2003), §§5.3, 7.5.

possible attack by NATO' and were not deployed with 'the intent to kill' people.[203] Whereas the Committee limited itself to answering questions of retroactive punishment and non-discrimination,[204] the ECtHR, in a related case, considered 'that recourse to anti-personnel mines and automatic-fire systems, in view of their automatic and indiscriminate effect, and the categorical nature of the border guards' orders to "annihilate border violators ... and protect the border at all costs"' flagrantly infringed fundamental human rights, including respect for and protection of the dignity and liberty of the person.[205]

Controlled, empty spaces are a central component of many imaginaries of autonomously secured boundaries.[206] The Korean DMZ, in particular, is presented in such narratives as 'the ideal location' for weapons like the SGR-A1 or the Super aEgis II because 'it is uninhabited', 'scrupulously guarded by thousands of soldiers on both sides',[207] and 'so heavily fortified that there are no civilians in it'.[208] Because 'any individuals that can physically enter' the weapon system's targeting range 'are reasonably presumed to be combatants',[209] the argument goes, it is unproblematic if a sentry system identifies any moving object with a human-body-shaped infrared heat-signature within its sensor range as a target. 'When you cross the line, you're automatically an enemy'.[210]

Essentially the same arguments used to advertise autonomous sentry systems in 'areas of exceptional conditions'[211] are also deployed in different settings, even if some supporters of autonomously secured boundaries concede that not all border areas are 'equally controlled environments' and caution against deploying autonomous systems in situations where the 'vast majority of border violators' are not 'military threats'.[212] For example, Israel has reportedly networked together remote-controlled machine guns, ground sensors, and drones along its border with Gaza to 'create 1500-meter deep "automated kill zones"'.[213] According to B'Tselem, the Israel Defense Forces classified substantial areas near the border fence as 'no-go zones' that people are prohibited from entering. Soldiers are allegedly allowed to open fire at anybody who enters that zone.[214] On the Gaza side, the boundaries of these 'no-go areas' are neither clearly marked nor fenced.[215]

B. THE DUTY TO INDIVIDUATE THE USE OF FORCE UNDER IHRL

Securing a zone or boundary with a sentry-AWS seems difficult to reconcile with legal precepts on the use of force for law enforcement purposes. For one, the deployment of such a system would have to be absolutely necessary in pursuit of a legitimate law enforcement objective. Under IHRL, the context within which the legality of the use of force is assessed includes both the specific circumstances of force application and 'all the surrounding circumstances, including ... the planning and control of the actions' and their regulation in abstract terms.[216] In the GDR cases, the ECtHR did not assess the 'automatic' and 'indiscriminate' effects of the weapons (ostensibly) deployed in isolation. It objected to the 'categorical' nature of the state *practice* for the use of force within which their deployment was embedded.[217] The Court considered that the border-policing *regime* 'clearly disregarded the need to preserve human life', and found that 'the deaths of the fugitives were in no sense the result of a use of force which was "absolutely necessary"' to secure a legitimate law enforcement aim as the practice did not protect anyone against unlawful violence, was not pursued in order to make a lawful arrest and did not serve to quell a riot or insurrection.[218] A 'general measure' preventing almost the entire population from leaving the GDR could not be necessary to protect

203 Ibid, §7.5. Similar arguments were advanced (equally unsuccessfully) by former political leaders of the GDR in *Streletz et al* (supra fn 201) and in ECtHR, *K.-H. W. v Germany*, App no 37201/97, Grand Chamber, Judgment, 22 March 2001.

204 *Klaus Dieter Baumgarten*, supra fn 202, §§9.3, 11.

205 *Streletz et al*, supra fn 201, §73.

206 See, e.g., A. Velez-Green, 'The Foreign Policy Essay: The South Korean Sentry – A "Killer Robot" to Prevent War', *Lawfare*, 1 March 2015, https://www.lawfareblog.com/foreign-policy-essay-south-korean-sentry%E2%80%94-killer-robot-prevent-war (envisaging the deployment of AWS for defensive use along borders, 'where civilians do not travel or can be prevented from traveling – where a controlled environment can reasonably be established').

207 Parkin, 'Killer Robots', supra fn 50.

208 Velez-Green, 'The Foreign Policy Essay', supra fn 206.

209 Ibid. Similarly, Lin et al, *Autonomous Military Robotics*, supra fn 72, p 77.

210 J. Kumagai, 'A Robotic Sentry For Korea's Demilitarized Zone', *IEEE Spectrum*, 1 March 2007, http://spectrum.ieee.org/robotics/military-robots/a-robotic-sentry-for-koreas-demilitarized-zone (citing Myung Ho Yoo, a principal research engineer at Samsung's Optics & Digital Imaging Division). In the same vein, Arkin, *Governing Lethal Behavior*, supra fn 81, p 93.

211 Arkin, *Governing Lethal Behavior*, supra fn 81, p 93.

212 Velez-Green, 'The Foreign Policy Essay', supra fn 206.

213 Shachtman, 'Robo-Snipers', supra fn 46; N. Shachtman, 'Israeli "Auto Kill Zone" Towers Locked and Loaded', *Wired*, 12 May 2008, https://www.wired.com/2008/12/israeli-auto-ki/. Israel has reportedly also automated its security along the Lebanese border ('Armour: Israeli Robots Roam the Earth', *StrategyPage*, 11 February 2011, https://www.strategypage.com/htmw/htarm/articles/20110211.aspx?comments=Y).

214 B'Tselem, 'Suspicion that Israel has Classified Areas near Gaza Perimeter Fence as «Killing Zones»', 27 February 2006, http://www.btselem.org/firearms/20060227_shooting_around_gaza_fence. According to the Palestinian Centre for Human Rights, persons have been attacked 'anywhere upto [*sic*] approximately 1.5 kilometres inside the border fence' ('PCHR-Gaza: Israeli Buffer Zone Policies Typically Enforced with Live Fire', *IMEMCnews*, 11 May 2015, http://imemc.org/article/71548/). Towers along the fence are reportedly equipped with 12.7mm machine guns, with a range of 2,000 meters, or with 7.62mm machine guns, with a range of 800 meters ('Armour: Israeli Robots Roam the Earth', supra fn 213).

215 In contrast to the inner German border fortifications, which were on the GDR's territory, the 'automated kill zones' along the Gaza-Israeli border are on Palestinian territory. See, e.g. 'Access and Closure: North Gaza, December 2014', Map, UN OCHA, June 2015, http://www.ochaopt.org/sites/default/files/North_Gaza1_july_2015.pdf.

216 *McCann et al*, supra fn 93, §150; ECtHR, *Andreou v Turkey*, App no 45653/99, Judgment, 27 October 2009, §§50, 51.

217 *Streletz et al*, supra fn 201, §73. In *Klaus Dieter Baumgarten* (supra fn 202, §7.2), Germany considered that the border guards' orders 'left no room for weighing the use of firearms against the principle of proportionality'.

218 *Streletz et al*, supra fn 201, §§96–97, 102.

the state's security.[219] Arguably, therefore, if an autonomous sentry system were deployed to prevent illegal immigration or unauthorized access to an airfield, port facility, warehouse, palace, pipeline, power plant or even an armoury, the aim pursued could hardly justify the institution of an automated kill zone and/or automated killing would be grossly disproportionate to the aim pursued. (The question of whether the situation would be different if an AWS were used to incapacitate, rather than to kill is explored further below.)

Automated kill zones are not, however, completely excluded under IHRL. In a case dealing with the emplacement by Turkish security forces of anti-personnel mines in a 'military security zone' around their station, the ECtHR did not challenge the government's argument that the mines were laid to protect the security forces. The Court accepted that the state had not deliberately sought to endanger the life of Erkan Erol, a boy who lost his leg after following his flock of sheep into the zone.[220] Instead, it referred to the state's duty to take appropriate steps to safeguard the lives of those within its jurisdiction against any real and immediate risk of which it has or ought to have knowledge. In the Court's assessment, the authorities had failed to take all security measures necessary to remove the risk of injury or death, and had thereby violated Erkan Erol's right to life.[221]

Spatial limitations on automated killing are clearly among the security measures necessary to remove the risk of injury or death, lest '[a]ll moments and all places' become 'potentially explosive traps, haunted by the possibility of killing'.[222] Yet, automated killing at the GDR's border was geographically bounded, the limits of 'exclusion zones' were clearly communicated to the population and regulatory and architectural measures were taken to prevent people from entering the area, and the practice was still found to be in breach of the obligation to respect human rights.[223] Similarly, the mined area at issue in *Paşa and Erkan Erol* was fenced, warning signs were installed and the inhabitants of the nearby village were informed about the danger. The Court nevertheless deemed these measures 'clearly insufficient' to prevent the entry of innocent civilians, including children, into the mined area, especially as it was located on the village's grazing land. This suggests that if a weapon is used whose parameters of a valid target are so broad that potentially lethal effects can be triggered by the presence of *anyone* entering the area (anyone and anything exerting a pressure of a set value and above in the case of a pressure-activated mine), the state *knowingly* exposes *everyone susceptible of entering the area* to a real and immediate risk to life.[224] The expectation on state agents using such a weapon system is, therefore, that they *effectively prevent* everyone who may fall within the system's target parameters, but who may *not* be legally killed, from entering the area. In practice, this expectation will prove difficult to fulfill[225] and raises the question of whether automated kill zones, that is, those that are autonomously secured, would be IHRL-compliant if the deployed weapon system were capable of reliably distinguishing between 'lawful targets' and 'innocent civilians'.

As the right to life is inherent in every person and not only in 'innocent civilians', it is not enough to stipulate that potentially lethal force may be used to prevent a certain category of people from illegally crossing a boundary (such as persons suspected or convicted of a violent crime approaching a police station or fleeing a high-security detention facility). To be legal, the use of lethal force must also be justified *in the concrete circumstances prevailing at the time*. It must be objectively reasonable for the state agent using force to believe that the person poses an *imminent threat of death or serious injury*. This condition holds even if refraining from the use of force may result in the person evading capture.[226] Cases dealing with incidents along the UN monitored 'buffer zone' in Cyprus[227] indicate that neither the failure of an individual illegally present in an exclusion zone to obey a warning,[228] nor

219 All human rights bodies dealing with the GDR border regime had regard to the fact that '[t]he killings took place in the context of a system which effectively denied to the population ... the right freely to leave one's own country' (*Klaus Dieter Baumgarten*, supra fn 202, §9.4; *Streletz et al*, supra fn 201, §63), exposing to mortal danger everyone who wanted to realize their right to freedom of movement. This speaks to the proportionality of the measure. In light of this, the fact that areas in the Gaza Strip deemed dangerous to access (up to 1500 meters from the border fence) comprise about 35% of the land that is suitable for farming impacts the legality of the border security regime put in place by Israel (B'Tselem, 'No-go Zones Near Gaza Strip Perimeter Fence', 1 January 2011, http://www.btselem.org/gaza_strip/forbidden_zones?level=1). Consider, in contrast, ECtHR, *Kallis and Androulla Panayi v Turkey*, App no 45388/99, Judgment, 27 October 2009, §63 (a Cypriot border guard voluntarily breached the ceasefire line to greet his Turkish counterpart); ECtHR, *Solomou et al v Turkey*, App no 36832/97, Judgment, 24 June 2008, §48 (a demonstrator voluntarily crossed the ceasefire line to climb a flagpole).

220 ECtHR, *Paşa and Erkan Erol v Turkey*, App no 51358/99, Judgment, 12 December 2006, §32.

221 Ibid, §§30–31, 37–38 (originally formulated in French as a requirement to 'prendre toutes les mesures afin d'empêcher la pénétration de civils innocents à cet endroit'). This positive duty exists irrespective of whether the state created the risk to life, but weighs all the more heavily if state agents contributed to its emergence. See, in particular, ECtHR, *Albekov et al v Russia*, App no 68216/01, Judgment, 9 October 2008, §§80–90 (where the Russian authorities denied having laid mines near a village but because they were aware of their existence the ECtHR considered that they had a positive obligation under Art 2 of the ECHR to locate and deactivate the mines, or failing this, to mark and 'seal off the area' so as to 'prevent anybody from entering it freely' and to 'comprehensively warn' the villagers of the location of the mines and the risks involved).

222 Bolton, 'From Minefields to Minespace', supra fn 87, 44.

223 *Streletz et al*, supra fn 201, §73; *K.-H. W.*, supra fn 203, §67.

224 The jurisprudential value of this case is somewhat diminished by the Court's failure to identify the correct international legal standards on the use of mines and the marking and fencing of mine fields. Instead, the Court cited a treaty among the applicable law that did not exist at the time of the events.

225 Although rare, North Korean defectors and, on one occasion, a South Korean civilian, have crossed the DMZ, for instance. See 'North Korea Soldier Walks over DMZ and Defects', *BBC News*, 15 June 2015, http://www.bbc.com/news/world-asia-33130382; A. C. Archive, 'South Korean "Defector" Wanted for Assault', *Asian Correspondent*, 28 October 2009, https://asiancorrespondent.com/2009/10/south-korean-defector-wanted-for-assault/.

226 *McCann et al*, supra fn 93, §200; ECtHR, *Makaratzis v Greece*, App no 50385/99, Grand Chamber, Judgment, 20 December 2004, §66; ECtHR, *Kakoulli v Turkey*, App no 38595/97, Judgment, 22 November 2005, §108; Inter-American Court of Human Rights (IACtHR), *Nadege Dorzema et al v Dominican Republic*, Judgment (Merits, Reparations and Cost), Series C no 251, 24 October 2012, §85.

227 The buffer zone and ceasefire line separating Greek Cypriote from Turkish Cypriote communities since 1974 is between 20 meters and 7 kilometers wide and contains minefields (UNSC Res 2300, 26 July 2016; 'UNFICYP Background', UN Department of Peacekeeping Operations, http://www.un.org/en/peacekeeping/missions/unficyp/background.shtml).

228 *Kakoulli*, supra fn 226, §116; *Kallis and Androulla Panayi*, supra fn 219, §62.

their (suspected) possession of a weapon[229] or wearing of an 'enemy' uniform[230] releases the authorities from assessing whether lethal force is absolutely necessary and strictly proportionate *in every individual case*.

The technology involved plays a role in this regard: when the use of an autonomous sentry system removes threats to life and limb of human security agents, it also removes a key justification for the recourse to deadly force.[231] Even in situations where the protection of life could, as a last resort, justify recourse to lethal force, human agents are nevertheless expected to retain 'the ability to assess all parameters and to organise their actions *carefully* with a view to *minimising* a risk of deprivation of life or bodily harm'.[232] Given the 'inherent need to make constant adjustments in a complex world',[233] putting in place a system where the potentially harmed *individual* is not considered in real time but, rather, factored into a predetermined process cannot be reconciled with the state's duty to 'strictly control and limit the circumstances in which a person may be deprived of their life'.[234] Lieblich and Benvenisti compellingly argue that the 'pre-binding' of targeting decisions inherent in the use of AWS contradicts the duty to 'give "due respect" to individuals by considering the effects of a specific act on individuals, in light of the prevailing circumstances'.[235] A requirement for human agents involved in the use of an AWS to remain constantly and actively (personally) engaged in every individual application of force is thus inherent in the duty to preserve human life. The need to individuate the use of force under IHRL severely limits the scope for the lawful use of an AWS for law enforcement purposes.[236]

This conclusion holds also for the use of force for law enforcement purposes during an armed conflict. Consider, for example, an AWS installed around a prisoner of war camp that would target any prisoner attempting to escape. The use of such a system would be unlawful under IHRL and IHL because it would apply *categorically* an 'extreme measure' that should be reserved for extraordinary circumstances.[237] According to an authoritative commentary on the Geneva Convention for the protection of prisoners of war, even when there is justification for opening fire and all the required material conditions have been met, such as when a prisoner crosses an established 'death line', '[f]ire may not be opened *automatically*'.[238] Considering the potential for riots within a camp to escalate to a situation that could ultimately make recourse to 'weapons of war' necessary, the same commentary notes that the Detaining Power must 'keep *a close watch* on the situation' to avoid such serious developments.[239] An escalation of force procedure is good practice in this and similar situations.

C. THE SCOPE FOR CATEGORICAL KILLING UNDER IHL

In contrast to law enforcement, under a conduct of hostilities paradigm, it can be permissible to target people not because their conduct poses an imminent threat of death, but on the basis of their legal status, that is, their (imputed) membership in a category of people who may be made the object of attack, such as 'combatants'.[240] IHL allows for 'categorical' (generic, corporate) killing in this sense and the vocabulary of proponents of autonomous sentry systems and automated kill zones betrays their assumption that IHL is, indeed, the primary legal frame of reference.[241]

One situation presumably governed by the law on hostilities would be the use of an autonomous sentry system to secure the perimeter around a military base in an area of active hostilities. The debate on the legality of AWS under IHL focuses on the question of whether, in such a situation, an AWS would be capable of directing force only at (human-body-shaped) objects that are legal targets *in the circumstances of an attack*. As noted previously, an attack would need to be sufficiently bounded to allow for the meaningful application of the rules on targeting. If an AWS, whose

229 *Kakoulli*, supra fn 226, §115.

230 *Kallis and Androulla Panayi*, supra fn 219, §60. See also Parkin, 'Killer Robots', supra fn 50 (reporting that DoDAAM engineers are hoping to develop systems 'able to computationally identify the type of enemy based on their uniform').

231 Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 370 (one of the problems presented by computer algorithms that determine when force will be released 'is that they do so in advance, on the basis of hypotheticals, while there is no true and pressing emergency rendering such a far-reaching decision unavoidable').

232 *Kakoulli*, supra fn 226, §114 (emphasis added); Principle 5(b), BPUFF.

233 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81, p 271.

234 HRCttee, *General Comment no 6: Article 6 (Right to Life)*, 30 April 1982, §3.

235 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81, p 271.

236 Left open is the possibility of a targeting system that autonomously determines the moment of force release, such as in the hostage scenario envisaged in Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 358: 'an AWS could conceivably be programmed, based on facial recognition, to release deadly force against a hostage-taker who is exposed for a split second, a situation in which a human sniper could be too slow to react, in a complex situation where the human mind cannot process all of the information in good time'. In such a scenario, the system functions autonomously within very narrow spatio-temporal boundaries and under constant, active human supervision, so as to ensure that *in the particular circumstances*, the system can be guaranteed to target the hostage-taker, but not others who (are made to) carry a weapon or have similar features or mannerisms (e.g. siblings or parents).

237 Art 42, 1949 Geneva Convention Relative to the Treatment of Prisoners of War (GC III) qualifies '[t]he use of weapons against prisoners of war, *especially* against those who are escaping or attempting to escape' as 'an extreme measure, which shall always be preceded by warnings appropriate to the circumstances' (emphasis added).

238 J. S. Pictet (ed), *Commentary: III Geneva Convention Relative to the Treatment of Prisoners of War*, ICRC, 1960, 2nd reprint, 2006, p 247.

239 Ibid, p 248 (emphasis added).

240 To what extent IHL limits status-based or spatially unbounded killing is subject to debate. See Melzer, *Targeted Killing in International Law*, supra fn 92, pp 397–399, for an argument on military necessity in its restrictive sense. Limitations can also be inferred from other central institutions of IHL, such as the protection of combatants *hors de combat*, necessitating an opportunity for surrender (Sparrow, 'Twenty Seconds to Comply', supra fn 172); protective presumptions to operate in case of doubt, e.g. about whether a person is a civilian (Art 50(1), AP I) or about the civilian character of an object 'normally dedicated to civilian purposes' (Art 52(3), AP I)) (Schmitt and Thurnher, '"Out of the Loop"', supra fn 19, 263, considering that AWS with 'adjustable doubt thresholds' would be an adequate solution); or respect for functional protection, as in the case of the medical mission.

241 See, e.g., K. Anderson and M. Waxman, 'Law and Ethics for Robot Soldiers', 176 *Policy Review* (1 December 2012), http://www.hoover.org/research/law-and-ethics-robot-soldiers, (considering '[a]lethal sentry robotdesigned for perimeter protection, able to detect shapes and motions, and combined with computational technologies to analyze and differentiate enemy threats from friendly or innocuous objects — and shoot at the hostiles').

sensor and weapons range is appropriately restricted, operates independently only for a limited period of time, objects that enter its sensor range and which are by their 'nature' military objectives (e.g. objects with the infrared signature and shape of enemy tanks) can arguably be treated as lawful targets of attack because their 'effective contribution' to enemy military action and the 'definite military advantage' of their destruction can be presumed *in these circumstances*.[242] However, the wider the spatial and temporal scope within which a system targets independently, or the broader the parameters of a valid target, the more problematic these presumptions become,[243] especially if human-body-shaped objects fall within its target parameters.[244]

It is contentious, however, whether 'autonomous targeting' (the use of an AWS to detect, select and engage targets without human intervention) is a permissible implementation of IHL rules on attack, notably, of the obligation to take 'constant care' to spare the civilian population by taking 'all feasible precautions' in attack, including by cancelling or suspending an attack 'if it becomes apparent that the objective is not a military one' or if 'the attack may be expected to cause civilian harm which would be excessive in relation to the concrete and direct military advantage anticipated'.[245] Even though the feasibility of precautionary measures and the proportionality of an attack are to be assessed in good faith by a 'reasonable military commander',[246] in conduct of hostilities-centered discourses, the permissibility of AWS use tends to be tied to an AWS' hypothetical capability to produce fewer wrongful casualties in the aggregate, compared to a 'human controlled system',[247] and it is sometimes argued that human involvement in the decision-making process that is temporally and geographically remote from the moment and location where violence is administered can still be adequate. Schmitt and Thurnher, for example, envisage the use of algorithms that allow an AWS to adjust its 'base maximum collateral damage threshold' and contemplate 'algorithms that can … precisely meter doubt' (a concept that, they concede, is framed in terms of 'human reasonableness').[248] IHL lends itself to such an orientation because a certain proportion of wrongfully killed persons is tolerated, whether as incidental casualties resulting from an attack on a lawful military objective or as the result of misidentification despite precautionary measures, and because IHL is silent about how key targeting rules are to be proceduralized.

Others consider that the rules of IHL preclude the removal of human agents from targeting decisions about specific attacks. Lieblich and Benvenisti argue that, as it cannot be determined in advance what is 'reasonable' or 'feasible' in any given situation, IHL notions of reasonableness and feasibility demand that 'the possibility of making adjustments' is left open. In their view, allowing final targeting decisions to be made by an AWS based on pre-programmed algorithms that cannot be altered in real time if circumstances require is unlawful in light of the duty to take *constant care* and thereby exercise continuous human discretion.[249] Similarly, Margulies derives a concept of 'dynamic diligence' from the rule on precautions. In his reading, a duty to exercise 'dynamic diligence' would not call for *ex ante* authorization of AWS targeting decisions, but would demand spatio-temporal limitations on an AWS' independent operation, coupled with 'frequent, periodic assessment and, where necessary, adjustment of AWS inputs, outputs, and interface with human service members'.[250] Although the precise procedural demands of IHL assessments and determinations remain nebulous in many regards, they demand *timely* adjustments and appear to assume capabilities of sensing, as well as capabilities of making sense,[251] suggesting that human agents using an AWS for the conduct of hostilities would need to exercise active and constant, in the sense of continuous or at least frequent, periodic, human control over individual attacks.

It bears repeating that even if there is scope for categorical targeting under IHL (that would not amount to an arbitrary deprivation of life), positive obligations

242 Art 52(2), AP I provides that '[i]n so far as objects are concerned, military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage.'

243 Note that proposals to list categories of legitimate objects of attack (as envisaged in, e.g., Art 7, 1956 ICRC Draft Rules for the Limitation of the Dangers incurred by the Civilian Population in Time of War) were rejected in the 1970s, in favour of a definitional approach that demands a contextual analysis in light of military necessity. In spite of this, proponents of AWS suggest that in order 'to program distinction' one could start with 'fixed lists of lawful targets' (Anderson and Waxman, 'Law and Ethics for Robot Soldiers', supra fn 241).

244 The precarious transition from being targetable to being protected against attack demands proximate human involvement. See, e.g., DPH Guidance, supra fn 106, pp 41–42 (noting that 'in determining whether a particular conduct amounts to direct participation in hostilities, due consideration must be given to *the circumstances prevailing at the relevant time and place*' (emphasis added); Sparrow, 'Twenty Seconds to Comply', supra fn 172 (stating that the nature of the signals used to indicate surrender is contextual and requires the ability to interpret and identify human intentions. Retaining the opportunity to surrender, he argues, requires keeping AWS on a 'tight leash').

245 Art 57(1) and (2)(b), AP I.

246 'Feasibility is an issue of reasonableness' (Schmitt and Thurnher, '"Out of the Loop"', supra fn 19, 261). See also G. Noll, 'Analogy at War: Proportionality, Equality and the Law of Targeting', 43 *Netherlands Yearbook of International Law* (2012) 205–230.

247 Schmitt and Thurnher, '"Out of the Loop"', supra fn 19, 261. See also Velez-Green, 'The Foreign Policy Essay', supra fn 206 (acknowledging that 'tragically', 'the initial non-combatants engaged by the machine' might not be saved 'as the operator would likely be unable to foresee the wrongful targeting and preemptively terminate the engagement', but suggesting that a sentry-AWS would be acceptable as long as the risk of 'wrongful targeting of non-combatants' would not do 'more harm than good').

248 Schmitt and Thurnher, '"Out of the Loop"', supra fn 19, 256–257, 263. See also M. Sassòli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified', 90 *International Law Studies* (2014), 322, 336 (assuming that machines act according to algorithms and, therefore, according to a plan established by humans). For a critique of this premise, see Suchman and Weber, 'Human-Machine Autonomies', supra fn 73, pp 85, 92; Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 371.

249 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81, p 270. This orientation does not mean that the use of a cruise missile is illegal, as long as the context of its use (the attack), and, thus, the space and time of independent functioning are sufficiently bounded and under human control.

250 Margulies, 'Making Autonomous Weapons Accountable', supra fn 85, pp 19, 22. See also US DoD, *Law of War Manual*, supra fn 9, s 6.5.9.3, p 330 (acknowledging that the obligation to take feasible precautions 'may be more significant' when a person uses 'weapon systems with more sophisticated autonomous functions', including 'monitoring the operation of the weapon system').

251 Noll, 'Analogy at War', supra fn 246, 223.

under IHRL continue to apply.[252] What operational steps are called for in terms of a state's positive obligation to safeguard life will vary in relation to the state's jurisdiction or exercise of authority, power or control over individuals or the area where the violence takes place. It is relevant to the assessment that modern weapon technologies offer unprecedented surveillance capabilities, and produce large amounts of data. In the Gaza border area, for example, control centres are reported to constantly video-record the Sentry Tech area of coverage to enable records of engagement.[253] The deployment of such technologies should, arguably, entail heightened expectations on states in terms of their positive obligations.[254] For instance, a state using an sentry-AWS could not invoke the removal or remoteness of human agents from the selection of specific targets to justify its failure to give effective advance warning of an attack that may affect the civilian population, or to recognize that a target was not a lawful one when these circumstances and the lack of awareness are the result of the state's own failure to take constant care in the conduct of its operations.[255]

D. 'NON-LETHAL' AUTONOMOUS INTERCEPTION

A key selling point for sentry systems that are not primarily or solely conceived for military combat appears to be their capability to '[d]etect ... an intruder or suspicious activities'[256] and apprehend a target. Such systems tend to be equipped with weapons branded as 'non-lethal'. The MDARS and the SCOUT, another patrol robot, are advertised for their 'interception'[257] and 'on-the-spot detainment of identified intruders',[258] or their ability to 'catch an escaped prisoner'.[259] The Guardium is described as being capable of 'suppress[ing] suspicious elements' and 'hold[ing] them back until manned security forces arrive', but it can also 'use various forceful means to eliminate the threat'.[260] The hyperbole of advertisers aside, the question can be asked whether the use of a sentry-AWS to intercept and incapacitate would be subject to lesser legal restraints than a potentially lethal one.

Whether the use of a sentry-AWS would conform to IHRL in a concrete situation depends on its legal basis, finality, necessity, proportionality and other factors, including, critically, its effects on the target and bystanders. 'Intercepting' people, for example to ascertain their identity, would constitute an interference with their right to freedom of movement.[261] Temporary restrictions on freedom of movement can be justified if these are 'provided by law' and 'necessary to protect national security' or 'public order', among other permissible purposes.[262] Importantly, though, detaining a person for even a 'very short duration' can amount to a deprivation of liberty.[263] Whereas the HRCttee considers that 'deprivation of liberty involves more severe restriction of motion within a narrower space than mere interference with liberty of movement',[264] for the ECtHR, the difference between a restriction on liberty of movement and a deprivation of liberty is 'merely one of degree and intensity, and not one of nature or substance'.[265] The distinction matters because a state assumes different and more extensive responsibilities vis-à-vis persons that it detains, compared to persons whose freedom of movement it has temporarily restricted. It is also worth noting that, to the extent that states detain persons not in relation to the prosecution for or prevention of a specific crime, but on the basis that they pose a security threat, the burden is on the state 'to show that

252 Although the ECtHR is ambiguous about whether it assesses the use of mines in *Paşa and Erkan Erol* (supra fn 220) and in *Albekov et al* (supra fn 221) as part of the conduct of hostilities, it is noteworthy that the Court found violations of the respective states' positive obligation to protect life in respect of situations that are widely regarded as NIACs (albeit not by the states involved).

253 Hughes, 'IDF Deploys Sentry Tech on Gaza Border', supra fn 52. In spite of the persistent video-recording of the Sentry Tech area of coverage, the Israeli army has so far 'refuse[d] to say how many Palestinians have been killed' by the system (J. Cook, 'Israel's Video Game Killing Technology', supra fn 63). See, e.g., A. Waked, 'Palestinians: 1 dead, 4 injured from IDF fire in Gaza', *Ynetnews*, 1 March 2010, http://www.ynetnews.com/articles/0,7340,L-3856218,00.html (reporting an incident involving Sentry Tech where it was 'unclear whether the casualties [were] farmers or gunmen').

254 Rosén, 'Extremely Stealthy and Incredibly Close', supra fn 130, 124–125.

255 A similar argument can be made under IHL. See, e.g., Boothby, 'Some Legal Challenges Posed by Remote Attack', supra fn 191, 584–585 (finding it unsatisfactory to argue 'that the absence of a human being from the autonomous aspect of the decision-making process renders the performance of these precautionary duties impractical' and thus 'militarily non-feasible'. At the same time, however, he considers that 'depending on the pattern of life in the relevant area, it may be possible to comply with the evaluative precautionary rules at the mission planning stage). Consider also ICRC Commentary on AP I, supra fn 170, §2221, p 686 (referring to the need for an attacker to 'observe' the context within which an attack is to take place, and pointing out that if direct observation is not possible due to the remoteness of the attacker, 'even greater caution is required').

256 Chun and Papanikolopoulos, 'Robot Surveillance and Security', supra fn 38, p 1606.

257 'Scout detected and confronted an intruder trying to gain unauthorized access to the flightline. After the intruder refused to obey commands issued by the controller, he was disabled with a pepper spray system mounted on Scout' (T. D. Erazo, 'Robotic Warriors Display Capabilities', *U.S. Air Force News*, 25 June 2004, http://www.af.mil/News/ArticleDisplay/tabid/223/Article/136631/robotic-warriors-display-capabilities.aspx).

258 General Dynamics Robotic Systems, 'MDARS', supra fn 62.

259 K. Grifantini, 'Modeling Sneaky Robots', *MIT Technology Review*, 20 May 2009, https://www.technologyreview.com/s/413544/modeling-sneaky-robots/.

260 'Enguard! Introducing the Guardium UGV', supra fn 59. Allegedly, it has 'already detained one fence-crosser' (B. Sweetman, 'Robot Sentry Patrols Borders', *Defense Technology International*, 1 October 2009, http://aviationweek.com/awin/robot-sentry-patrols-borders). An autonomous system can be advertised for use in a wide variety of situations, blurring the line between perimeter security and crowd control. According to TechnoRobot, the manufacturer of RiotBot, 'scenarios that have been studied for [the] development' of the robot include 'riot control', 'civil order', 'jails and prisons', as well as 'area denial' and 'urban warfare'. (TechnoRobot, 'Riotbot's Applications', http://www.technorobot.eu/en/riotbot.htm). The RiotBot is equipped with ammunition that 'combines a frangible plastic sphere with a concentrated load of PAVA (Capsaicin II) powder, an active and potent ingredient of pepper spray' and can deliver over 700 rounds per minute. (TechnoRobot, 'Rapidly Deployable Remotely Operated Less-Lethal Support Robots', *armytechnology.com*, http://www.army-technology.com/contractors/unmanned_vehicles/technorobot/).

261 Art 12, ICCPR; Art 12, African Charter on Human and Peoples' Rights (AfCHPR); Art 22, American Convention on Human Rights (AmCHR); Art 2, Protocol 4 to the ECHR.

262 Art 12(3), ICCPR.

263 Art 9, ICCPR; Art 6, AfCHPR; Art 7, AmCHR; Art 5, ECHR.

264 HRCtte, *General Comment no 35: Article 9 (Liberty and security of person)*, 16 December 2014, §5.

265 ECtHR, *Gillan and Quinton v The United Kingdom*, App no 4158/05, Judgment, 12 January 2010, §§56–57 (considering that the 'element of coercion' is 'indicative of a deprivation of liberty', but making no determination on Art 5).

the *individual* poses such a threat'.[266] Likewise, only the '*individualized* likelihood of absconding, a danger of crimes against others or a risk of acts against national security' can justify continued detention of a person who entered the country illegally. Such decisions cannot be based on 'a mandatory rule for a broad *category*'.[267]

Furthermore, use of force that does not result in or is not intended to cause death can nevertheless fall within the ambit of the right to life.[268] An AWS can conceivably harm through kinetic energy (causing a projectile to hit or penetrate a target or by creating blast overpressure through a detonation, for example) or by other means, including the diffusion of chemical substances or the direction of electromagnetic energy. In the framework of the CCW, discussions on AWS are formally limited to the use (and development) of 'lethal' AWS[269] in the context of an armed conflict (or more accurately, their use as a means of warfare).[270] The categorization of weapons into 'lethal' and 'non-'or 'less lethal' obscures that the effects of a weapon are never solely a function of its design, but also depend on its use and the vulnerabilities of those affected by it.[271]

Whereas a focus on 'lethal' AWS draws attention to the risk of death directly resulting from AWS use, physical harm short of death, severe mental suffering and material damage are also humanitarian and human rights concerns. Serious physical injury or severe mental trauma can amount to inhuman or degrading treatment.[272] Civilian objects are specifically protected against direct attack under IHL,[273] and in addition to direct effects on people (or of 'anti-personnel' AWS), the use of an AWS against an object (or of an 'anti-materiel' AWS) can have indirect adverse effects on human health, including death, for example to people in the vicinity of an object of attack.[274]

266 Ibid, §§15 (emphasis added).

267 Ibid,§18 (emphasis added).

268 E.g., ECtHR, *Nakayev v Russia*, App no 29846/05, Judgment, 21 June 2011, §58.

269 Final Report, Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects, UN doc CCW/MSP/2015/9, 27 January 2016, §§5, 35.

270 Preamble and Art 1, CCW. Originally, the scope of the CCW was limited to IACs. An amendment to Art 1 adopted on 21 December 2001 expands its scope to NIACs. The use of weapons in 'riots, isolated and sporadic acts of violence, and other acts of a similar nature' are excluded from the CCW's ambit, as is the use of weapons for law enforcement purposes during an armed conflict.

271 Weapons termed 'non-lethal' are capable of causing and have in fact caused death in some circumstances. See, e.g., A. Dymond-Bass and N. Corney, 'The Use of "Less-lethal" Weapons in Law Enforcement', in S. Casey-Maslen (ed), *Weapons Under International Human Rights Law*, Cambridge University Press, 2014, p 33.

272 As recognized in ECtHR, *Abdullah Yaşa et al v Turkey*, App no 44827/08, Judgment, 16 July 2013 (serious injury from the use of tear-gas grenades during demonstrations) and in ECtHR, *Benzer v Turkey*, App no 23502/06, Judgment, 12 November 2013 (witnessing the killing of close relatives and wanton destruction of applicants' houses by airstrikes on a village).

273 Arts 52-56, AP I.

274 In any case, in military parlance, 'kill' or 'lethal' does not necessarily imply death in a medical sense. The notion of 'lethal area', for instance, is used in weaponeering to express the effectiveness of a particular weapon against a specific target (M. R. Driels, *Weaponeering: Conventional Weapon System Effectiveness*, American Institute of Aeronautics and Astronautics, 2nd edn, 2013, pp 283–284).

The use of a sentry-AWS to intercept and potentially incapacitate people could, thus, interfere with the right to freedom of movement, and in particular circumstances, constitute a measure of a 'coercive and restrictive nature', amounting to a deprivation of liberty.[275] It could even fall within the ambit of the right to life or the prohibition of cruel, inhuman or degrading treatment. The availability of equipment that allows for a differentiated use of force can help minimize the risk of injury and damage,[276] but this does not release the state from the duty to 'carefully evaluate' its deployment and to 'carefully control' its use.[277] The type of weapon used affects whether the use of force is deemed proportionate and necessary in a particular situation, but case law indicates that it is not decisive whether it is characterized as use of a 'non-lethal incapacitating weapon', 'lethal force', 'potentially lethal force' or as force that is not usually fatal.[278] Whether such use in a particular situation is lawful will depend on a range of factors, including the type, duration, effects and manner of implementation of the measure. Without going into detail, the limited argument advanced here is that because the legal duties flowing from the same security measure involving an AWS can vary considerably depending on individual circumstances, compliance with IHRL requires essentially the same type of *individuated* human control in the use of an autonomous sentry system, irrespective of whether it is equipped with weapons branded as 'non-lethal' and intended to 'intercept' rather than 'eliminate'.

In times of war or other public emergency, states can derogate from their obligations under the right to liberty of movement and the right to liberty and security of person,[279] and under IHL states are permitted to take quite intrusive 'measures of control', includ-

275 ECtHR, *Austin et al v The United Kingdom*, App nos 39692/09, 40713/09 and 41008/09, Grand Chamber, Judgment, 15 March 2012, §§60, 64–69 (noting that it cannot be excluded that the use of containment and crowd-control techniques such as 'kettling' could, in particular circumstances, give rise to an unjustified deprivation of liberty); ECtHR, *Gahramanov v Azerbaijan*, App no 26291/06, Decision, 15 October 2013, §§39, 41 (finding that airport border control holding a passenger is not a deprivation of liberty if the 'detention' does not exceed 'the time strictly necessary to comply with relevant formalities'. In this case, the intervention of a human agent permitted the timely correction of a database error that caused the 'detention').

276 Principle 2, BPUFF.

277 Principle 3, BPUFF.

278 In ECtHR, *Scavuzzo-Hager et al v Switzerland*, App no 41773/98, Judgment, 7 February 2006, §§56–63, the Court grappled with the differentiation between force that is not 'in itself fatal' but is nevertheless 'susceptible to lead to death'. Whereas the ECtHR tends to describe the use of firearms or explosive weapons as 'lethal' or 'potentially lethal' use of force (ECtHR, *Finogenov et al v Russia*, App nos 18299/03 and 27311/03, Judgment, 20 December 2011, §232: considering that bombs and missiles are 'supposed to kill'), it has recognized that riot-control agents like tear gas bear a 'risk of causing serious injury … or indeed of killing someone' (*Abdullah Yaşa et al*, §§42–43). In ECtHR, *Ataykaya v Turkey*, App no 50275/08, Judgment, 22 July 2014, §46, the Court described the use of tear-gas grenades that resulted in the death of a demonstrator both as 'lethal force' and as 'potentially lethal force' in the same paragraph.

279 For example, in June 2015, the Ukraine notified its derogation from Art 2 of Protocol 4 to 'temporarily restrict freedom of movement and the right to private life', including by granting 'powers to military and civil administrations' to enforce curfews and 'to temporarily restrict or prohibit the movement of vehicles and pedestrians on the streets, roads and terrain areas' (Permanent Representation of Ukraine to the Council of Europe, Note Verbale, 5 June 2015, https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=2833408&SecMode=1&DocId=2278178&Usage=2).

ing assigned residence and internment.[280] Such measures are not, in principle, arbitrary under IHRL to the extent that they are authorized and regulated by and comply with IHL.[281] Even during an armed conflict, though, recourse to such measures can, generally speaking, only be justified, exceptionally, by imperative security reasons and decisions must be made *on an individual basis*, in regular judicial or administrative proceedings. Internment decisions must be individuated in NIACs, and there is a strong presumption that this is also the case in relation to persons protected under the Fourth Geneva Convention (civilians in IACs).[282] In contrast, IHL allows for the 'categorical internment' of persons who by virtue of their 'status' are protected as prisoners of war under the Third Geneva Convention (combatants falling into the power of the enemy in IACs).[283] This opens up the theoretical possibility of using an AWS to intercept and detain persons in an IAC but, in practice, procedural safeguards under both IHRL and IHL that protect individuals against arbitrariness would seem to severely limit the scope of such a measure.[284]

E. ALGORITHMIC TARGET CONSTRUCTION: A THREAT TO HUMAN RIGHTS AND HUMAN DIGNITY

Requirements and constraints on the use of AWS have thus far been derived mainly from the need to adequately situate legal evaluations (to individuate them under IHRL) with a view to assessing the legality of the use of force in terms of its outcome. Yet, IHRL also places demands on decision-making processes, including in terms of how and why persons can lawfully be made the target of security measures. This last section of the study elucidates aspects of algorithm-based target construction that threaten human dignity, as well as the right to privacy, the right not to be discriminated against and not to be subjected to cruel, inhuman or degrading treatment and the right to an effective remedy. From this perspective, human intervention in the use of an AWS serves as a procedural safeguard to uphold human dignity and human rights.

280 E.g. Arts 35 (right to leave the territory); 41 (assigned residence and internment); 49 (deportations, transfers and evacuations in occupied territory); 78 (assigned residence or interment in occupied territory), 1949 Geneva Convention Relative to the Protection of Civilian Persons in Time of War (GC IV).

281 HRCtte, *General Comment no 35*, supra fn 264, §64.

282 See, e.g., N. Melzer, *International Humanitarian Law: A Comprehensive Introduction*, ICRC, 2016, p 191 (noting that 'the mere fact that a person is an enemy national cannot be regarded as a security threat *automatically* justifying internment without completely defeating the idea of tailoring security measures to the requirements of each *individual* case and reserving internment for the most serious cases' (emphasis added)). In contrast, J. S. Pictet (ed), *Commentary IV Geneva Convention Relative to the Protection of Civilian Persons in Time of War*, ICRC, 1960, 2nd reprint, 2006, Art 41(1), p 256 (reporting that an explicit mention of a requirement to take decisions on internment individually was discussed at length and finally rejected by the drafters, 'on the ground that there might be situations – a threat of invasion for example – which would force a government to act without delay to prevent hostile acts, and to take measures against certain categories without always finding it possible to consider individual cases'. According to the Commentary, this was compensated by procedural safeguards against arbitrariness).

283 Art 21 (internment of prisoners of war), GC III.

284 See ECtHR, *Hassan v The United Kingdom*, App no 29750/09, Grand Chamber, Judgment, 16 September 2014, §106. For a discussion, see J. Pejic, 'Procedural Principles and Safeguards for Internment / Administrative Detention in Armed Conflict and Other Situations of Violence', 87 *IRRC* 858 (2005) 375–392.

1. SURVEILLANCE

Imaginaries of autonomous targeting take shape against the backdrop of intrusive, secret, surveillance systems put in place in recent years in the name of counter-terrorism/violent extremism.[285] Developments in surveillance practices and information technologies are generating ever larger amounts of digitized data to which statistical, data analysis or AI techniques can be applied, with minimum human intervention.[286] Using an AWS capable of detecting individuals or objects matching certain criteria and tracking them entails surveillance of people's habits of everyday life, places of residence, movements, activities, social relationships and social environments frequented by them. Such use would, thus, likely involve the automatic processing of personal data[287] with the potential to undermine key data protection principles.[288] Whereas an autonomous sentry system used to analyse patterns within a delimited zone would expose people to surveillance and the risk of being targeted during their presence in that zone, an AWS could conceivably be given access to surveillance infrastructures and data held anywhere in the world, enabling persistent surveillance, and, possibly, attacks on targets within a wide

285 S.Carlo, 'The Snooper's Charter Passed into Law this Week – Say Goodbye to Your Privacy', *The Independent*, 19 November 2016, http://www.independent.co.uk/voices/snoopers-charter-theresa-may-online-privacy-investigatory-powers-act-a7426461.html.

286 J.-M. Dinant , C. Lazaro, Y. Poullet, N. Lefever and A Rouvroy, *Application of Convention 108 to the Profiling Mechanism: Some Ideas for the Future Work of the Consultative Committee*, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, 11 January 2008, p 5, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680684 0b9.

287 Art 2(a), 1981 Council of Europe (CoE) Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (CETS 108) defines 'personal data' as 'any information relating to an identified or identifiable individual'. The definition also covers 'sets of data which are geographically distributed and are brought together via computer links for purposes of processing' (Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 28 January 1981, §30). It also includes situations where it is possible to identify an individual through the combination of different types of data, such as physical, physiological, genetic, economic, or a combination of data on age, sex, occupation, geolocation, family status, etc. Whereas an individual is not considered 'identifiable' if his or her identification would require 'unreasonable time, effort or resources', technological developments such as growing processing power, 'may change what constitutes 'unreasonable time, effort or other resources' (Draft Explanatory Report to Draft Modernized CETS 108, §§16, 18, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806b6ec2). Art 2(b), Council of Europe Draft Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data (Draft Modernized CETS 108), as of September 2016, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a616c,defines 'data processing' as 'any operation or set of operations which is performed on personal data, such as the collection, storage, preservation, alteration, retrieval, disclosure, making available, erasure, or destruction of, or the carrying out of logical and/or arithmetical operations on such data'.

288 For example, data mining for targeting risks infringing the principle of purpose specification and limitation if it entails subsequent processing of data that modifies the purposes originally justifying the data collection. See Art 5(4)(b), Draft Modernized CETS 108 and Draft Explanatory Report to Draft Modernized CETS 108, supra fn 287, §§46–47; European Union Agency for Fundamental Rights (FRA) and the Council of Europe, *Handbook on European Data Protection Law*, 2014, pp 68-70, http://www.echr.coe.int/Documents/Handbook_data_protection_ENG.pdf.

area or in a geographically unbounded manner (the 'hunter-killer scenario').[289]

In recent years, leading human rights actors have expressed growing concern about the negative impact that the blanket interception of communications and mass collection of personal data, including extraterritorially, can have on the enjoyment of human rights.[290] The systematic collection and storing of data by security services or other public authorities relating to an individual's life constitute an interference with the right to respect for privacy,[291] a fundamental human right that serves to safeguard human dignity.[292] The detection of serious crime or prevention of terrorist acts, or more broadly, the safeguarding of national security or defence can justify interference with the right to privacy but strict conditions apply.

When processing personal data, human dignity requires that adequate safeguards be put in place in order for individuals 'not to be treated as mere objects'.[293] What safeguards will prove effective in a particular case will depend on the nature, scope and duration of the measure, the grounds required for ordering it, the authorities competent to permit, carry out and supervise it, and the kind of remedy provided by law. *Secret* surveillance bears a particular risk of abuse and arbitrariness.[294] Even in the name of national security, powers of secret surveillance are tolerable only in so far as they are strictly necessary for safeguarding the democratic institutions.[295] So-called 'strategic', automated or large-scale interception that effectively results

289 Consider R. Leheny, 'DARPA's Urban Operations Programs', DARPATech, 9–11 August 2005, p 38 http://archive.darpa.mil/DARPATech2005/presentations/diro/leheny.pdf; 'Combat Zones that See', in US DoD, *Department of Defense Fiscal Year (FY) 2005 Budget Estimates*, February 2004, pp 170–171, http://www.darpa.mil/attachments/(2G12)%20Global%20Nav%20-%20About%20Us%20-%20Budget%20-%20Budget%20Entries%20-%20FY2005%20(Approved).pdf; Anderson and Waxman, 'Law and Ethics for Robot Soldiers', supra fn 241 (envisaging '[t]iny surveillance robots equipped with facial recognition technology' to identify specific persons suspected of terrorist acts, and noting that '[i]t is not a large step to weaponize such systems and then perhaps go the next step to allow them to act autonomously'). For a critical discussion of such imaginaries, see S. Graham, 'Technologies of Exception: Urban Warfare and US Military Technoscience', Conference Lecture at the Symposium 'Arxipelago of Exceptions - Sovereignties of Extraterritoriality', CCCB 10–11 November 2005, available at publicspace.org, http://www.publicspace.org/es/texto-biblioteca/eng/b022-technologies-of-exception-urban-warfare-and-us-military-technoscience. See also D. Wilson, 'Military Surveillance', in K. Ball, K. D. Haggerty and D. Lyon (eds), *Routledge Handbook of Surveillance Studies*, Routledge, 2012, p 274 (cautioning that 'the long-standing tendency of military technologies to migrate into civilian application should alert us to the possibility of technophiliac dreams [of an automated surveillance-military killing machine] taking root in domestic contexts).

290 UNGA Res 68/167, 21 January 2014; *The Right to Privacy in the Digital Age*, Report of the Office of the United Nations High Commissioner for Human Rights, UN doc A/HRC/27/37, 30 June 2014; HRCttee, *Concluding Observations on the Fourth Periodic Report of the United States of America*, UN doc CCPR /C/USA/CO/4, 23 April 2014, §22. Several complaints about mass surveillance are pending before the ECtHR.

291 Art 17, ICCPR; Art 11 AmCHR; Art 8, ECHR.

292 Preamble, Draft Modernized CETS 108, supra fn 287.

293 Draft Explanatory Report to Draft Modernized CETS 108, supra fn 287, §9.

294 ECtHR, *Klass et al v Germany*, App no 5029/71, Judgment (Plenary), 6 September 1978, §49 ('an unlimited discretion to subject persons within their jurisdiction to secret surveillance' would risk 'undermining or even destroying democracy on the ground of defending it').

295 Ibid, §42. See also ECtHR, *Roman Zakharov v Russia*, App no 47143/06, Grand Chamber, Judgment, 4 December 2015, §§231–232; ECtHR, *Rotaru v Romania*, Grand Chamber, Judgment, 4 May 2000, §47 and Concurring Opinion of Judge Wildhaber (joined by six other judges) (noting that 'States do not enjoy unlimited discretion to subject individuals to secret surveillance or a system of secret files').

in 'the unlimited surveillance of a large number of citizens' is not acceptable.[296]

In a recent case concerning national legislation requiring the retention of personal data by telecommunication providers and access of the authorities to this data, the European Union (EU) Court of Justice made it clear that however fundamental 'an objective of general interest' the fight against terrorism or organized crime may be, it 'cannot in itself justify that national legislation providing for the *general and indiscriminate* retention of all traffic and location data should be considered to be necessary for the purposes of that fight'.[297] Without analysing this important case in detail, suffice it to underline here that the algorithmic construction of targets of security measures builds on practices that are already identified as deeply problematic from a human rights angle. Also, relying on autonomous targeting would, arguably, further the trend toward justifying the necessity of security measures on a relatively abstract level, *ex ante*, and in a *generalized* manner, rather than *differentiating* and *limiting* such measures to a *particular* time period, geographical area or group of persons likely to pose a security threat.[298]

2. SORTING PEOPLE

It has been suggested that AWS 'may be programmed in part based on "pattern of life analysis" of the target area'.[299] In such a scenario, an AWS would detect individuals (or groups) who possess certain attributes that are believed to bear positive *statistical correlations* to particular kinds of conduct, such as involvement in terrorism or participation in hostilities, but whose identities need not be known. The mining of massive datasets offers numerous possibilities for categorizing individuals on the basis of some observable characteristics so as to infer other characteristics that are not observable with a view to taking individual decisions relating to them or *predicting* their attitude or behaviour.[300] This constitutes 'profiling'.[301]

296 ECtHR, *Szabó and Vissy v Hungary*, App no 37138/14, Judgment, 12 January 2016, §§56, 67–69, 71, 73. The Court considers it 'a natural consequence' of 'present-day terrorism that governments resort to cutting-edge technologies' in pre-empting terrorist attacks, including 'the massive monitoring of communications' and 'automated and systemic data collection'. But for the Court, 'so-called strategic, large-scale interception' is 'a matter of serious concern' in that it paves the way for 'the unlimited surveillance of a large number of citizens'. In the Court's view, 'it would defy the purpose of government efforts to keep terrorism at bay … if the terrorist threat were paradoxically substituted for by a perceived threat of unfettered executive power intruding into citizens' private spheres by virtue of uncontrolled yet far-reaching surveillance techniques and prerogatives'. See also Court of Justice of the European Union, *Digital Rights Ireland and Others v Minister for Communications, Marine and Natural Resources and Others*, Case nos C-293/12 and C-594/12, Grand Chamber, Judgment, 8 April 2014, ECLI:EU:C:2014:238.

297 Court of Justice of the European Union, *Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others*, Case nos C203/15 and C698/15, Grand Chamber, Judgment, 21 December 2016, ECLI:EU:C:2016:970, §103 (emphasis added).

298 Ibid, §§106–110.

299 Schmitt and Thurnher, '"Out of the Loop"', supra fn 19, 268.

300 The aim is 'to aid a decision modifying a course of action while ensuring that choices already made are applied automatically and more effectively' (Dinant et al, *Application of Convention 108 to the Profiling Mechanism*, supra fn 286, pp 3, 11); FRA, *Towards More Effective Policing, Understanding and Preventing Discriminatory Ethnic Profiling: A Guide*, 2010, p 9, https://fra.europa.eu/sites/default/files/fra_uploads/1133-Guide-ethnic-profiling_EN.pdf.

301 CoE Committee of Ministers Recommendation CM/Rec(2010)13 on the Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling, 23 November 2010 (CoE CM/Rec(2010)13), §§1(d) and (e).

Profiling places people in categories, usually without their knowledge, on the basis of information pertaining to them *not as individuals*, but because of their (imputed) membership in that category. The categorization of individuals based on correlations and inferences always entails a certain error rate. The very presupposition that 'relevant circumstances can be rendered algorithmically, and still adequately, as "patterns of life"' is of course questionable.[302] The approach can also be criticized for its dehumanizing quality in that 'it tends to reduce the person to the profile generated by automated processes which are liable to be used as a basis for decision-making',[303] This is 'one of the most acute dangers of profiling'[304] as it facilitates the translation of bodies into targets for security measures, including measures involving armed force.[305]

When public authorities base decisions on the processing of personal data by complex, opaque algorithms this risks adversely affecting human dignity and encroaching on a range of freedoms and rights, including economic and social rights. Whereas any law and order determination should be based on an individual's personal conduct,[306] algorithmic processes put individuals under categorical suspicion due to their (imputed) membership in a category perceived by the authorities, rightly or wrongly, as being dangerous.[307] 'Individuals therefore face the risk of a prediction of their future behaviour that is nothing to do with them and is no more than a forecast based on the previous behaviour of other individuals whom they do not know.'[308] Subjecting people to law enforcement measures on this basis is unlawful.[309]

Moreover, profiling involving 'special categories of data' (sensitive data) such as genetic or biometric data, or personal data revealing information about racial or ethnic origin, political opinions, religious or other beliefs, health or sexual life,[310] bears a particularly high risk of negatively affecting human dignity and of exposing individuals to discrimination.[311] Accounts of unfair and stigmatizing impacts of algorithm-based decisions abound, from credit scoring by private actors, to predictive policing and the compilation of 'no-fly lists' by executive authorities, and risk assessments in the criminal justice system.[312] It is illegal to subject persons to surveillance, identity checks or the use of force, and thus interfere with their rights and freedoms, solely (or mainly) on grounds of their race, national or ethnic origin, gender, sex or religion.[313] To be lawful, such measures must rely on additional factors that give 'reasonable grounds' for targeting a particular person, and, in practice, they must not mainly or more negatively impact one particular group compared to another in a manner that cannot objectively be justified as proportionate and necessary to achieve a legitimate objective.[314]

Targeting security measures based on broad-brush racial, ethnic, religious and national origin stereotypes in order to identify potential threats and vulnerabilities has been harshly criticized in the counter-terrorism context.[315] It is doubtful that the severe in-

302 Suchman and Weber, 'Human-Machine Autonomies', supra fn 73, p 91.

303 Dinant et al, *Application of Convention 108 to the Profiling Mechanism*, supra fn 286, p 6.

304 Ibid.

305 On this aspect, in relation to armed drones, see, in particular, Wall and Monahan, 'Surveillance and Violence from Afar', supra fn 44, 239–254.

306 Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance, UN doc A/HRC/29/46, 20 April 2015, §39.

307 E.g. ECtHR, *Ostendorf v Germany*, App no 15598/08, Judgment, 7 March 2013, §66 (finding a deprivation of liberty for preventive purposes lawful, not least because the police 'had *not* based their findings on entries on the applicant in a police database on persons prepared to use violence in the context of sports events' (Ibid, §80, emphasis added)).

308 Dinant et al, *Application of Convention 108 to the Profiling Mechanism*, supra fn 286, p 34.

309 ECtHR, *Shimovolos v Russia*, App no 30194/09, Judgment, 21 June 2011, §§54–56 (finding that the subjection of a human rights activist to law enforcement measures on the basis that his name was included in a 'Surveillance Database' set up by Russian authorities to facilitate discovery of 'potential extremists' was impermissible as it amounted to 'a policy of general prevention' directed against an individual or a category of individuals who are perceived by the authorities, rightly or wrongly, as being dangerous or having propensity to unlawful acts. The Court underlined that mere 'membership' in a human rights organization cannot form a sufficient basis for a suspicion justifying the arrest of an individual).

310 Art 6, Draft Modernized CETS 108, supra fn 287.

311 Preamble, CoE CM/Rec(2010)13, supra fn 301. The right to equality and non-discrimination prohibits discrimination in law and in fact in any field regulated and protected by public authorities (among other legal bases, Arts 2 and 26, ICCPR; Arts 1 and 24, AmCHR; Art 2, AfCHPR; Art 14, ECHR; HRCtte, *General Comment no 18: Non-Discrimination*, 10 November 1989, §12). The principle of non-discrimination in data processing is enshrined, e.g., in Article 1, CETS 108. See also UN doc A/HRC/29/46, supra fn 306, §22 (expressing concern about racial and ethnic profiling being institutionalized through risk profiling software).

312 M. Stroud, 'The Minority Report: Chicago's New Police Computer Predicts Crimes, but is it Racist?', *The Verge*, 19 February 2014, http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist; L. Dormehl, 'Algorithms Are Great and All, But They Can Also Ruin Lives', *Wired*, 19 November 2014, https://www.wired.com/2014/11/algorithms-great-can-also-ruin-lives/; S. Ackermann, 'No-Fly List Uses "Predictive Assessments" Instead of Hard Evidence, US Admits', *The Guardian*, 10 August 2015, https://www.theguardian.com/us-news/2015/aug/10/us-no-fly-list-predictive-assessments; A. M. Barry-Jester, B. Casselman and D. Goldstein, 'The New Science of Sentencing', *The Marshall Project*, 8 April 2015, https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.mku7Zs9Ym.

313 See e.g., UN doc A/HRC/29/46, supra fn 306 (looking at ethnic and racial profiling by law enforcement agencies); Report of the Special Rapporteur on Contemporary Forms ofRacism, Racial Discrimination, Xenophobia and Related Intolerance on the Manifestations of Defamation of Religions, and in Particular on the Ongoing Serious Implications of Islamophobia, for the Enjoyment of all Rights by their Followers, UN doc A/HRC/15/53, 12 July 2010.

314 States must put in place effective safeguards against discrimination in purpose and effect. The processing of sensitive data requires additional safeguards and profiling through automated decision making that results in discrimination is prohibited (Art 6, Draft Modernized CETS 108, supra fn 287; Principle 2.4, CoE Committee of Ministers, Recommendation no R(87)15 Regulating the Use of Personal Data in the Police Sector, 17 September 1987; Art 3.11, CoE CM/Rec(2010)13, supra fn 301; Art 11, Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016).

315 Report of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms While Countering Terrorism, UN doc A/HRC/4/26, 29 January 2007, §34 ('profiling based on stereotypical assumptions that persons of a certain "race", national or ethnic origin or religion are particularly likely to commit crime may lead to practices that are incompatible with the principle of non-discrimination'); ECtHR, *10 Human Rights Organizations et al v The United Kingdom*, App no 24960/15 (pending, lodged on 20 May 2015, concerning the NSA's secret mass surveillance programs, complaining of a violation of Art 14 (non-discrimination) in conjunction with Arts 8 and 10 in respect of foreign nationals); American Civil Liberties Union, Letter to Attorney General Holder, 20 October 2011, https://www.aclu.org/files/assets/aclu_letter_to_ag_re_rm_102011_0.pdf and 'Latif, et al. v. Lynch, et al. – ACLU Challenge to Government No Fly List', 20 August 2015, https://www.aclu.org/cases/latif-et-al-v-lynch-et-al-aclu-challenge-government-no-fly-list?redirect=cases/latif-et-al-v-holder-et-al-aclu-challenge-government-no-fly-list.

terference with the rights of everyone affected by algorithmic targeting can ever be justified as necessary and proportionate to safeguard democratic institutions.[316]

As noted previously, there is scope for 'categorical targeting' within a conduct of hostilities framework but the principle of non-discrimination continues to apply in armed conflict. Adverse distinction based on race, sex, religion, national origin or similar criteria is prohibited.[317] Remarkably, under Additional Protocol II to the Geneva Conventions, applicable in certain NIACs, the prohibition on adverse distinction also applies to persons directly participating in hostilities.[318] The legality of drone strikes purportedly connected to an armed conflict, where targets are selected on the basis of a number of observable, behavioural or other 'signatures', has been challenged on the grounds that categories used do not map exactly onto the legal definitions of persons or objects that may legally be made the object of attack,[319] and for being based on insufficient evidence that the targets exhibited characteristics or behaviour justifying attack.[320] It has also been pointed out that the categories underpinning 'signature strikes' are gendered and racialized in a manner that exposes men of a particular age group and religion in certain geographic areas to a disproportionate risk of extrajudicial killing.[321]

As for the human rights challenges linked to surveillance, the risks posed by profiling are not limited to AWS. However, algorithm-based decisions reflect structural

316 In 2006, the German Federal Constitutional Court held that a secret, automated data-mining measure could not be justified by reference to a general terrorist or security threat or a situation of increased tension. To be compatible with the fundamental right to informational self-determination, a right derived from human dignity and the right to the free development of one's personality, such a measure could only be justified by a 'concrete threat' to national security or to a person's life, limb or freedom. The case concerned the automated processing of massive amounts of data by the German Federal Police in the wake of the terrorist attacks of 11 September 2001 with a view to identify terrorist 'sleepers' in Germany on the basis of such criteria as age, educational enrolment, faith, country of birth and nationality and sex of a person. The Court considered that the automated, computerized nature of the measure, allowing for the fast processing of large, complex datasets, contributed to the severity of the interference with the complainant's rights. Other aggravating factors included the types of personal data collected (including sensitive personal data), the linking of datasets, the potential for adverse consequences, including stigmatization and increased risk of becoming the subject of unjustified suspicion or investigation, the secrecy of the measure and its wide scope (affecting between 200,000 and 300,000 people). The requirement of a concrete threat does not demand an imminent or present threat, but it demands that the facts in a concrete case indicate with sufficient probability that a concrete threat will materialize (German Federal Constitutional Court, Beschluss des Ersten Senats, 4 April 2006, 1 BvR 518/02 - Rn. (1-184), http://www.bverfg.de/e/rs20060404_1bvr051802.html).

317 Common Art 3, GC I–IV; Art 75, AP I; Art 2(1) AP II.

318 S. Krähenmann, 'The Obligation Under International Law of the Foreign Fighter's State of Nationality or Habitual Residence, State of Transit and State of Destination', in A. de Guttry, F. Capone and C. Paulussen (eds), *Foreign Fighters Under International Law and Beyond*, T. M. C. Asser Press, 2016, p 256.

319 K. Benson, '"Kill 'em and Sort it Out Later!": Signature Drone Strikes and International Humanitarian Law', 27 *Pacific McGeorge Global Business & Development Law Journal* 1 (2014) 31–32. Benson identifies the 'inability to contextualize' as a source of disproportionate civilian harm (37).

320 See, e.g., K. J. Heller, '"One Hell of a Killing Machine": Signature Strikes and International Law', 11 *Journal of International Criminal Justice* 1 (2013) 1-22.

321 R. Acheson, R. Moyes and T. Nash, *Sex and Drone Strikes: Gender and Identity in Targeting and Casualty Analysis*, Reaching Critical Will and Article 36, October 2014, http://www.reachingcriticalwill.org/images/documents/Publications/sex-and-drone-strikes.pdf.

biases in society. AWS do not themselves make legal judgements or discriminate but when machines are made to learn from datasets, they reify existing patterns of discrimination.[322] Autonomous targeting threatens to perpetuate and has the potential to reinforce essentialist stereotypes that are already recognized as adversely affecting human dignity and equality in present practice.

3. CALCULATED BLINDNESS

Whereas conduct of hostilities-oriented narratives about AWS readily accept that '[n]o algorithmic system of any useful complexity will deliver perfectly accurate results',[323] these tend to obfuscate that to make autonomous targeting work, some tolerance of error needs to be defined. Put differently, 'probabilistic matching … *requires* the deliberate targeting of noncombatants as a statistically necessary function of the system'.[324] Calibrating sensor technology is, thus, not an ethically neutral act,[325] and on some level, 'wrongfully targeted' people are 'more than merely foreseen because they result from the meticulous programming of the device, which is a deliberate act that sits somewhere in a causal chain between volition and an outcome'.[326] There is a difference between human agents having to take into account the inaccuracy of unguided artillery when ordering shelling, and the acceptance, *ex ante*, of a percentage of false positives in autonomous targeting.[327]

Others have effectively critiqued autonomous targeting for precluding 'deliberative human intervention'[328] and the exercise of discretion.[329] Even in war, as Asaro explains, 'the authority to decide to initiate the use of lethal force cannot be legitimately delegated to an automated process' because human agents who use force are required 'to reflexively consider the implications of their actions, and to apply

322 K. Crawford, 'Artificial Intelligence's White Guy Problem', *The New York Times*, 25 June 2016, http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=1. Failing to recognize these structural biases bears a significant risk of regression as illustrated by X. Wu and X. Zhang, *Automated Inference on Criminality using Face Images*, last revised 21 November 2016, arXiv:1611.04135v2 [cs.CV] (proposing the automated prediction of criminality using face recognition technology, purportedly 'free of any biases').

323 E. Stoddart, 'A Surveillance of Care: Evaluating Surveillance Ethically', in K. Ball, K. D. Haggerty and D. Lyon (eds), *Routledge Handbook of Surveillance Studies*, Routledge, 2012, p 375.

324 M. S. Swiatek, 'Intending to Err: The Ethical Challenge of Lethal, Autonomous Systems', 14 *Ethics and Information Technology* 4 (2012) 241 (emphasis added).

325 Stoddart, 'A Surveillance of Care', supra fn 323, p 375.

326 Swiatek, 'Intending to Err', supra fn 324, 247.

327 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81, p 276-277.

328 Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 370; P. Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making', 94 *IRRC* 886 (2012) 695.

329 Lieblich and Benvenisti, 'The Obligation to Exercise Discretion', supra fn 81. Similarly, Suchman and Weber, in 'Human-Machine Autonomies', supra fn 73, p 92 (describing the world as 'an open horizon of potentially relevant circumstances', where relevance is continuously, socially constructed and cannot be adequately encoded *ex ante*).

compassion and judgement in an explicit appeal to their humanity'.[330] Subjecting individuals to measures of 'an automatic nature' without an explanation or limitation as to their scope or duration,[331] treating all people (indiscriminately) 'like objects',[332] and failing to carry out a fresh review taking account of individual circumstances is unlawful in many situations. It is particularly problematic when the consequences are irreversible.

In the context of judicial executions, for example, the mandatory application of the death penalty for particular categories of offences, precludes 'reasoned consideration of each individual case', including mitigating factors and different levels of criminal responsibility, and cannot be 'the subject of an effective review by a higher court'.[333] This has been found to produce arbitrary results and to amount to cruel, inhuman and degrading punishment.[334] Courts have rejected the categorical, automatic application of the death penalty as 'blind adherence to the letter of the law',[335] that treats people 'not as uniquely individual human beings, but as members of a faceless, undifferentiated mass',[336] that is 'degrading because it strips the convicted person of all dignity and treats him or her as an object to be eliminated by the state',[337] and that is irreconcilable with 'the essential respect for the dignity of the individual'.[338]

330 Asaro, 'On Banning Autonomous Weapon Systems', supra fn 328, 689, 700; See also M. Koskenniemi, 'Faith, Identity, and the Killing of the Innocent: International Lawyers and Nuclear Weapons', 10 *Leiden Journal of International Law* (*LJIL*) (1997) 159–160 (noting that '[d]iscretion and "evaluation", even error and misjudgment are part of the law, however much it is dressed in the voice of universal reason'); Chamayou, *Drone Theory*, supra fn 1, p 218 ('in redefining "ethical" as conforming mechanically to rules, it is reduced to being synonymous with the most lobotomized discipline or docility'. Ruling out the very possibility of disobedience comes 'at the cost of simultaneously suppressing the principal source of infralegal limitation to armed violence: the critical conscience of its agents').

331 ECtHR, *Battista v Italy*, App no 43978/09, Judgment, 2 December 2014, §§47–48 (automatic withdrawal of passport); ECtHR, *Stamose v Bulgaria*, App no 29713/05, Judgment, 27 November 2012, §§33–36 (automatic imposition of a travel ban). In contrast, e.g., ECtHR, *Landvreugd v The Netherlands*, App no 37331/97, Judgment, 4 June 2002, §70 (finding no disproportionate restriction on freedom of movement as a result of time-limited designations of certain areas of Amsterdam as 'emergency areas' because the authorities had ascertained the individual would not suffer undue hardship).

332 E.g. *Gillan and Quinton*, supra fn 265, Joint Dissenting Opinion of Judges Tulkens, Spielmann and Garlicki, §10 (arguing that 'kettling' was 'applied indiscriminately' as 'all people who happened to be at Oxford Circus at around 2p.m. were treated *like objects* and were force to remain there as long as the police had not solved other problems around the city' (emphasis added)).

333 IACommHR, *Cases 12.023 (Desmond McKenzie), 12.044 (Andrew Downer y Alphonso Tracey), 12.107 (Carl Baker), 12.126 (Dwight Fletcher) and 12.146 (Anthony Rose) v Jamaica*, Report no 41/00, OEA/Ser.L/V/II.106 doc 3 rev at 918 (1999), §§194–196. See also HRCttee, *Mr Rawle Kennedy v Trinidad and Tobago*, Decision (Comm no 845/1998), UN doc CCPR/C/74/D/845/1998 (2002), §7.3.

334 A. Priddy and M. Mattirolo, 'The Mandatory Death Penalty Under International Law', Geneva Academy of International Humanitarian Law and Human Rights, Final Draft, February 2013, unpublished, p 3. The focus on the mandatory application of the death penalty is not meant to suggest that the death penalty as such conforms with IHRL.

335 Supreme Court of India, *Mithu et al v State of Punjab et al*, Judgment, 7 April 1983, 2 SC R 690, Headnotes, §(ix).

336 US Supreme Court, *Woodson v North Carolina*, Judgment, 2 July 1976, 428 US 280 (1976), Syllabus, §(d).

337 Constitutional Court of the Republic of South Africa, *T. Makwanyane and M. Mchunu v The State*, Case no CCT/3/94, Judgment, 6 June 1995, §26.

338 *Cases 12.023 (Desmond McKenzie), 12.044 (Andrew Downer y Alphonso Tracey), 12.107 (Carl Baker), 12.126 (Dwight Fletcher) and 12.146 (Anthony Rose)*, supra fn 333, §203.

Even if *extra*judicial killing is not about 'rendering justice',[339] 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.[340]

4. PROCESS MATTERS

To police the problems raised by automated decisions, human oversight and involvement are essential. It has long been recognized that every individual should have the right to ascertain 'in an intelligible form', whether, and if so, what personal data is stored in automatic data files, and for what purposes, to know who controls that data and to request rectification or elimination.[341] The opportunity to be heard 'at a meaningful time and in a meaningful manner' to challenge an automated decision is a key element of due process.[342] Recent regulatory efforts in the European context recognize the right not to be subject to a decision significantly affecting a person based solely on an automated processing of data without an opportunity to challenge such a decision, including one's assignment to a particular category.[343] An EU Directive concerning the processing of personal data for police and criminal justice purposes, which takes effect in 2018, establishes a presumption that subjecting persons to such automated decisions is prohibited, unless authorized by law and only if suitable safeguards are provided. These safeguards include the right to obtain *human intervention*, in particular, to receive an explanation of the decision reached or to challenge the decision.[344]

339 Sassòli, 'Autonomous Weapons and International Humanitarian Law', supra fn 248, 332.

340 On the tension between the dominant 'techno-fantasy' of achieving omniscience trough all-seeing technologies and the creation of blindness to critical differentiations, see Wilson, 'Military Surveillance', supra fn 289, p 274; Chamayou, *Drone Theory*, supra fn 1, p 42 (explaining in relation to 'signature strikes' that 'identification' predicated on an analysis of behaviour patterns is not individual but generic); Wall and Monahan, 'Surveillance and Violence from Afar', supra fn 44, 240 (describing drones as forms of surveillance that accord with the precepts of categorical suspicion and social sorting that 'force homogenization upon difference').

341 HRCttee, *General Comment no 16: Article 17 (Right to Privacy)*, 8 April 1988, §10.

342 US District Court for the District of Oregon, *Latif et al v Holder et al*, Case no 3:10-cv-00750-BR, Opinion and Order, 24 June 2014 (striking down the existing redress procedure against inclusion on a no-fly list as unconstitutional).

343 Art 8(1)(a), Draft Modernized CETS 108, supra fn 287; Preamble, §38 and Art 11, Directive (EU) 2016/680, supra fn 314.

344 Preamble, §38 and Art 11, Directive (EU) 2016/680, supra fn 314 (emphasis added). Note, however, that the Directive does not apply to national security activities or to agencies or units dealing with national security issues (Preamble, §14). CETS 108 has a broader scope of application but neither Art 8, Draft Modernized CETS 108 nor §73(a), Draft Explanatory Report to Draft Modernized CETS 108 (supra fn 287) refer explicitly to *human* intervention, and both the right not to be subjected to a decision based solely on an automated processing of data and the right 'to obtain, on request, knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her' can be restricted for the protection of national security or defence when this 'is provided for by law, respects the essence of the fundamental rights and freedoms and constitutes a necessary and proportionate measure in a democratic society' pursuant to Art 9(1)(a), Draft Modernized CETS 108. See also Dinant et al, *Application of Convention 108 to the Profiling Mechanism*, supra fn 286, p 34.

European law also recognizes the pressing importance of *knowing the reasoning* underlying decisions based on the processing of one's data,[345] 'in particular in cases involving the use of algorithms for automated-decision making including profiling'.[346] As mentioned earlier, however, 'it is becoming increasingly clear that human beings may not necessarily always be able to understand how (and possibly why) autonomous systems make decisions'.[347] According to Goodman and Flaxman, recognizing 'a right to explanation' could 'require a complete overhaul of standard and widely used algorithmic techniques',[348] as common supervised machine learning algorithms are not built with a concern for *causal* reasoning in mind.

This is a major worry from the perspective of states' procedural obligations under IHRL and the right to an effective remedy.[349] As the ECtHR has repeatedly explained, the legal prohibition of arbitrary killing by state agents would be ineffective in practice if there existed no procedure for reviewing the lawfulness of the use of force by the authorities. When individuals have been killed as a result of the use of force, there should be some form of effective, independent investigation.[350] This procedural obligation under the right to life continues to apply 'in difficult security conditions, including in a context of armed conflict'.[351] To be effective, an investigation, among other aspects, has to be *capable* of leading to a determination of whether the force used was or was not justified *in the circumstances*, and to the identification and punishment of those responsible.[352] Accordingly, human agents must be in a position to provide a concise, intelligible account of how input features relate to predictions and categorizations.[353] In human rights proceedings, the onus is on the state to provide sufficient details on its decision-making procedures to allow an independent assessment of the legality of the use of force and to assist victims and society at large in their quest for the truth.[354] It would seem to follow that using force by means of a technology that renders an investigation into resulting deaths *a priori* incapable of determining whether force was justified in *particular* circumstances violates the right to life.[355]

Concerns have been raised that autonomous targeting risks being effectively unchallengeable and outside of judicial supervision.[356] The challenges are not unlike those raised by secret surveillance. In that connection, the ECtHR has held that an individual can claim to be a victim of a violation of the ECHR occasioned by *the mere existence* of secret surveillance measures or *legislation permitting such measures* if two conditions are met:[357] one, the scope of the legislation is such that the person can 'possibly be affected', 'either because he or she belongs to a group of persons targeted' or because the legislation institutes a system where the communications of 'any person' can be intercepted; and, two, the domestic system does not afford an effective remedy to anyone who 'suspects that he or she was subjected to secret surveillance' without that person having to demonstrate the existence of a risk that surveillance was applied to them.[358] It is arguable by analogy that the very deployment of an AWS in an area would entitle people potentially falling within its target parameters and sensor and weapons range to the opportunity to challenge

345 Art 8(1)(c), Draft Modernized CETS 108, supra fn 287.

346 Draft Explanatory Report to Draft Modernized CETS 108, supra fn 287, §75.

347 Anderson et al, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems', supra fn 19, 394.

348 Goodman and Flaxman, 'European Union Regulations on Algorithmic Decision-Making', supra fn 75, p 1.

349 Explicit in, e.g., Art 2(3), ICCPR; Art 13, ECHR; Art 25, AmCHR.

350 *McCann et al*, supra fn 93, §161; *Al-Skeini*, supra fn 100, §163.

351 *Al-Skeini*, supra fn 100, §164. See also ECtHR, *Kaya v Turkey*, App no 22729/93, Judgment, 19 February 1998, §91.

352 *Jaloud*, supra fn 123, §166. Consider, e.g. ECtHR, *Isayeva v Russia*, App no 57950/00, Judgment, 24 February 2005, §§221–223 (finding that the investigation was not effective, and noting that it 'made surprisingly few attempts to find *an explanation* for these serious and credible allegations', denying the applicants 'any realistic possibility … to challenge the conclusions' of the authorities' account (emphasis added)).

353 In this vein, D. Keats Citron and F. Pasquale, 'The Scored Society: Due Process for Automated Predictions', 89 *Washington Law Review* (2014) 1–33 (detailing 'technological due process' requirements) https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1318/89WLR0001.pdf?sequence=1; N. Diakopulous, 'How to Hold Governments Accountable for the Algorithms They Use', *Slate*, 11 February 2016, http://www.slate.com/articles/technology/future_tense/2016/02/how_to_hold_governments_accountable_for_their_algorithms.html (proposing the introduction of a 'Freedom of Information Processing Act'). See also ECtHR, *Khodorkovskiy and Lebedev v Russia*, App nos 11082/06 and 13772/05, Judgment, 25 July 2013, §848 (finding a violation of Art 8 due to the failure to 'explain how' a plan was drawn up to distribute convicts among prisons and to describe the method or algorithm used). With respect to the conduct of hostilities, see Margulies, 'Making Autonomous Weapons Accountable', supra fn 85, p 23 (asserting the need to have recourse to correct errors, and for interpretability, and noting that 'nomination decisions by AWS should be interpretable and transparent, and that if a nomination is mistaken and raises questions about compliance with IHL principles, 'a state should be able to present a clear account of the AWS's calculation').

354 This concerns, notably, information about targeting decisions, including the criteria for selecting targets and precautions incorporated in such criteria. See, e.g., Report of the Detailed Findings of the Independent Commission of Inquiry Established Pursuant to Human Rights Council Resolution S-21/1, UN doc A/HRC/29/CRP.4, 24 June 2015, §§216–218; Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, 13 September 2013, UN doc A/68/382, §98.

355 If a government's failure to submit information or otherwise to provide a satisfactory and convincing explanation in a human rights proceeding prevents a court from reaching factual conclusions, the court can draw inferences in favour of the applicant (e.g. IACtHR, *Case of Velásquez Rodríguez v Honduras*, Judgment, 29 July 1988, Series C no 4, §§127-146). A failure to account for the fate of an individual can result in a violation of the right to life, including in cases where the involvement of state agents is disputed or cannot be established. See, e.g., ECtHR, *Varnava et al v Turkey*, App nos 16064/90, 16065/90, 16066/90, 16068/90, 16069/90, 16070/90, 16071/90, 16072/90 and 16073/90, Grand Chamber, Judgment, 18 September 2009, §§184–186, 191, 194 (failure to account for the whereabouts and fate of men who disappeared 'in life-threatening circumstances where the conduct of military operations was accompanied by widespread arrests and killings'); ECtHR, *Osmanoglu v Turkey*, App no 48804/99, Judgment, 24 January 2008, §§84, 92 (failure to take reasonable measures to prevent a real and immediate risk to the life of a man, including failure to launch an investigation into his disappearance). For an in-depth discussion of the duty to account, see Krähenmann, 'Positive Obligations in Human Rights Treaties', supra fn 95, pp 214–233.

356 HRW and IHRC, *Mind the Gap*, supra fn 114, pp 27–29.

357 *Roman Zakharov*, supra fn 295, §163; *Klass et al*, supra fn 294, §34.

358 *Roman Zakharov* , supra fn 295, §171. See also, e.g., ECtHR, *Colon v The Netherlands*, App no 49458/06, Decision, 15 May 2012, §60 (concerning the application of 'preventive search' orders in designated 'security risk areas' of Amsterdam, the Court recognized victim status in the absence of individual measures of implementation, if individuals 'run the risk of being directly affected' by the measure, that is, if they are 'required either to modify their conduct or risk being prosecuted, or if they are members of a class of people who risk being directly affected by the legislation').

the legality of its deployment, including extraterritorially.[359] The burden would be on the state using the AWS to demonstrate that any limitations on the enjoyment of rights serve a legitimate purpose and do not render the essence of the rights meaningless, and that the safeguards in place provide effective protection against unlawful, arbitrary, disproportionate or discriminatory interference.[360]

Finally, the GDR cases discussed above touch upon another issue at the heart of the present debate on AWS: the relationship and tension that can exist between positive law and justice[361] and the accountability of human agents involved, however remotely, in automated killing. All applicants in the GDR cases claimed that it had been impossible for them to foresee that they would one day be called to account in a criminal court. All review bodies rejected this argument on the grounds that respect for and protection of fundamental human rights and faith in the dignity and worth of human beings were already at the time general principles of law recognized by the community of nations, and which were reflected in domestic law.[362] Political leaders, the ECtHR considered, could not be ignorant of the international obligations entered into by their state or of the repeated international criticism of the security regime they put in place. Nor could they rely on laws and regulations that they themselves had put in place.[363]

6. CONCLUDING REMARKS

The advent of increasing autonomy in weapon systems poses new challenges to the international regulation of the use of force for the protection of the human person, and acerbates existing ones.

The use of force by means of an AWS, in pursuit of a legitimate law enforcement objective would expose anyone falling within the parameters of a valid target to a real and immediate risk to life. To safeguard life, the state deploying the AWS has a duty to take *all* measures necessary to *effectively prevent* anyone potentially

359 E.g. *Roman Zakharov*, supra fn 295, §§286–301; *Rotaru*, supra fn 295, §69. See also, Draft Explanatory Report to Draft Modernized CETS 108, supra fn 287, § 24 ('Any data processing carried out by a public sector entity falls directly within the jurisdiction of the Party, as it is the result of the Party's exercise of its jurisdiction').

360 E.g. *Roman Zakharov*, supra fn 295, § 284 ('it is for the Government to illustrate the practical effectiveness of the supervision arrangements'); *Szabó and Vissy*, supra fn 296, § 88.

361 In a decision of 24 October 1996, the German Federal Constitutional Court, relying on the so-called 'Radbruch Formula', held that in extraordinary cases, where positive law is intolerably inconsistent with justice, the principle of legal certainty may have to yield precedence to that of objective justice (cited in *Streletz et al*, supra fn 201, §22).

362 Ibid, §75; *K.-H. W.*, supra fn 203, §§56-57.

363 *Streletz et al*, supra fn 201, §103; *Klaus Dieter Baumgarten*, supra fn 202, § 4.2. Even security agents of a lower rank cannot 'show total, blind obedience to orders which flagrantly infringe recognized human rights, including the right to life' (*K.-H. W.*, supra fn 203, §75). Cf, however, Partly Dissenting Opinion of Judge Pellonpää, Joined by Judge Zupančič (differentiating between high-ranking officials and agents of a lower rank in terms of their responsibility).

falling within the system's target parameters, but who may not be legally killed, from entering the system's sensor and weapon range. Even when recourse to lethal force can in abstract terms be justified, it must also be absolutely necessary and strictly proportionate in a concrete situation. When there is no imminent threat to life or risk of serious injury, including due to the removal of state agents from the location where force is administered, recourse to lethal force cannot be justified as absolutely necessary. To comply with the requirement that lethal force be used only as a last resort whilst minimizing the risk of deprivation of life or bodily harm, human agents must be continuously and actively engaged in every instance of force application. Due to the need to individuate the use of force, the scope for autonomous targeting is extremely limited under IHRL.[364]

The requirement to place strict spatio-temporal limitations on the use of force by means of an AWS follows not only from the duty to safeguard life, but more broadly from the need to evaluate the legality of security measures, including those interfering with the rights to freedom of movement and to security and liberty of person, in the circumstances of *every particular case*. Compliance with IHRL requires essentially the same type of individuated human control in the use of an autonomous sentry system, irrespective of whether it is equipped with weapons branded as 'non-lethal' and intended to 'intercept' rather than 'eliminate'. Although there is limited scope for a more categorical approach to detention in IACs, even under IHL there is a presumption that decisions on detention must, in principle, be made on an individual basis. For all practical purposes, therefore, human state agents 'must remain *personally* in control of the actual delivery or release of force, in a manner capable of ensuring respect for the rights of *any particular individual*, as well as the general public',[365] both during and outside of armed conflict.

In relation to hostilities, where IHRL standards on the use of force are interpreted in light of IHL, there is some scope for categorical targeting, which allows broadening the context of evaluation to that of an attack (as a whole). However, attacks must remain sufficiently bounded in spatio-temporal terms to allow the application of legal rules by human agents. This includes, notably, the obligation to take all feasible precautions in attack, from which can be derived a requirement on human agents to retain control sufficient to recognize changing circumstances and to make adjustments in a timely manner. Arguably, this calls for active and constant, in the sense of continuous or at least frequent, periodic, human control over every individual attack. Human control over AWS during the conduct of hostilities must also safeguard the opportunity to shift to a law enforcement model when this becomes factually possible and, thus, legally mandated.

In addition to a duty to ensure that the outcome of a security measure comports with legal requirements, IHRL articulates demands on decision-making processes, including in terms of how and why persons may be targeted or killed. The algorith-

364 Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 362–366; HRW and IHRC, *Shaking the Foundations*, supra fn 22, pp 9–14.

365 AfCommHPR, *General Comment no 3*, supra fn 84, s E, §31. (emphasis added).

mic construction of targets draws on practices that are already considered deeply problematic from a human rights perspective, including secret mass surveillance, large-scale interception of personal data and algorithm-based profiling. The use of AWS is likely to sustain and even promote such practices, threatening human dignity, the right to privacy, the right not to be discriminated against and not to be subjected to cruel, inhuman or degrading treatment and the right to an effective remedy. Targeting based on 'patterns of life' analyses, for instance, places individuals under categorical suspicion, is blind to critical differences and stigmatizing in its effects. The automaticity and objectification inherent in target construction by means of an AWS, and the absence of deliberative human intervention is dehumanizing. To safeguard human dignity and human rights, human agents must remain involved in algorithmic targeting processes in a manner that enables them to explain the reasoning underlying algorithmic decisions in concrete circumstances. This is essential to ensuring the availability of an effective remedy, accountability for the use of force and for maintaining public confidence in states' adherence to the rule of law, in times of peace as well as war.

By focusing on IHRL requirements and constraints on the use of an AWS, this study challenges the appropriateness of existing IHL as the sole means of regulating AWS. An IHRL-oriented approach allows autonomous weapon technologies to be situated against the backdrop of practices of automated killing that are deeply contested and have been severely criticized by human rights bodies. It draws attention to the objectifying, dehumanizing and potentially discriminatory processes involved in autonomous targeting, rather than being concerned with its outcomes alone. It reminds us of the responsibilities we 'normally' expect states to assume vis-à-vis their own populations, and helps confine militarized rationalities and technologies to the exceptional and extraordinary. It counters the fabrication of irresponsibility,[366] and favours a precautionary orientation: whereas 'the core design of IHL is consistent with promoting, rather than restricting' purportedly value-neutral new technologies,[367] from an IHRL perspective, the introduction of technologies (of violence) that inhibit or diminish, rather than facilitate or promote a state's capacity to fulfill its human rights obligations and to safeguard human life and dignity, cannot be acceptable.[368]

366 Asaro, 'Determinism, Machine Agency, and Responsibility', supra fn 82, 292 ('The consequence of this is that we end up with organizations and systems that are increasingly designed for irresponsibility.'); Chamayou, *Drone Theory*, supra fn 1, p 211.

367 Kerr and Szilagyi, 'Evitable Conflicts, Inevitable Technologies?', supra fn 40, 31.

368 In this vein, Heyns, 'Human Rights and the Use of Autonomous Weapons Systems', supra fn 22, 374 ('If a state uses weapons systems that prima facie limit rights such as those listed above, the onus to show that it is justified under human rights law is thus clearly on the state … If there is doubt, for example, as to whether humans retain sufficient levels of control over the release of force not to implicate human dignity, such use of force should not be permissible.'); UN doc A/65/321, supra fn 12, §48 (recommending proactive steps to ensure that new technologies are optimized in terms of their capacity to *promote more effective compliance* with IHL and IHRL); AfCommHPR, *General Comment no 3*, supra fn 84, s F, §35 ('The use during hostilities of new weapons technologies such as remote controlled aircraft should only be envisaged if they *strengthen* the protection of the right to life of those affected').

The law is a 'crucial means by which the economy of violence is calculated and managed'.[369] Legal norms already regulate and limit algorithmic decision making and automated killing but new technologies and evolving security practices challenge the categories and disrupt the human–machine configurations around which the legal regulation of force is articulated. This generates controversies and uncertainties about the applicability and meaning of existing norms, thus diminishing existing law's capacity to serve as a guidepost. There is also the risk that '[e]stablished norms and rules of international law are preserved formally, but filled with a radically different meaning', for, accommodating a practice in legal terms 'means that international law itself is undergoing a transformation'.[370] An explicit, formal, legal requirement for the exercise of meaningful human control in the use of force can help safeguard human dignity and human rights.

Increasing autonomy in weapon systems is neither automatic nor inevitable. Inevitability is purposefully constructed by human agents.[371] It is an ethical question and a political act when human agents attribute agency to a technological device or system rather than to people.[372] This returns responsibility to us as representatives of institutions that deploy the technology, who are involved in its design, who use the equipment or, perhaps most significantly, who are subjected to its operation.[373]

369 E. Weizman, *The Least of All Possible Evils: Humanitarian Violence from Arendt to Gaza*, Verso, 2011, p 4.

370 S. Krasmann, 'Targeted Killing and its Law: On a Mutually Constitutive Relationship', 25 *LJIL* 3 (2012) 674; Kerr and Szilagyi, 'Evitable Conflicts, Inevitable Technologies?', supra fn 40, 28 ('law can be transformed by a collective omission or new practice … achieved through the introduction of a new technology that "forces" new practices').

371 Chamayou, *Drone Theory*, supra fn 1, p. 211. See also Suchman and Weber, 'Human-Machine Autonomies', supra fn 73, p 91 (on the conflation of the descriptive and the promissory); Kerr and Szilagyi, 'Evitable Conflicts, Inevitable Technologies?', supra fn 40, 25 ('what technology makes possible has the power to generate in our minds what may later be perceived of as necessary').

372 Koskenniemi, 'Faith, Identity, and the Killing of the Innocent', supra fn 330, 160 (the dominant juridical discourse 'perpetuates the illusion of the existence of a privileged (legal) rationality that is able to resolve any political conflict without becoming political itself. Its bureaucratic attachment to legal technique allows the abdication of personal responsibility for anything that can be supported by this technique – and anything can').

373 Stoddart, 'A Surveillance of Care', supra fn 323, p 375.

The Geneva Academy provides post-graduate education, conducts academic legal research and policy studies, and organizes training courses and expert meetings. We concentrate on branches of international law that relate to situations of armed conflict, protracted violence, and protection of human rights.

Established in 2007 by the Faculty of Law of the University of Geneva and the Graduate Institute of International and Development Studies, the Geneva Academy has acquired a global reputation for excellent teaching and research, and it attracts students of high quality to its master's and training programmes. Our more than 660 graduates are employed around the world, promoting and protecting International humanitarian law (IHL) and human rights (HR) in governments, NGOs, international organizations and academic institutions. The Geneva Academy thus contributes to the dissemination of legal knowledge in these crucial sectors.

Our scientific research focuses on clarifying IHL, strengthening human rights protection, and developing the areas of complementarity between IHL and international human rights law. In these areas, the Geneva Academy makes a specific contribution to policy development and debate, in government and among scholars and practitioners.

The Geneva Academy is a cosmopolitan community located in the heart of Geneva, an international city and humanitarian hub. Through close interaction with international organizations, NGOs, experts, and governments, we actively participate in global discussions of IHL, HR, international criminal law, and transitional justice.

978-29701003-3-1

The Geneva Academy
of International Humanitarian Law
and Human Rights

Villa Moynier
Rue de Lausanne 120B
CP 1063 - 1211 Geneva 1 - Switzerland
Phone: +41 (22) 908 44 83
Email: info@geneva-academy.ch
www.geneva-academy.ch