

1

# Autonomous weapons systems: living a dignified life and dying a dignified death

CHRISTOF HEYNS

## Introduction

The ever-increasing power of computers is arguably one of the defining characteristics of our time. Computers affect almost all aspects of our lives and have become an integral part not only of our world but also of our very identity as human beings. They offer major advantages and pose serious threats. One of the main challenges of our era is how to respond to this development: to make sure computers enhance and do not undermine human objectives. <sup>1</sup>

The imposition of force by one individual against another has always been an intensely personal affair – a human being was physically present at the point of the release of force and took the decision that it would be done. It is inherently a highly controversial issue because of the intrusion on people's bodies and even lives. Ethical and legal norms have developed

This contribution overlaps with, and draws on, a number of earlier ones by the same author: a report to the Human Rights Council, Christof Heyns, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Doc. A/HRC/23/47, 9 April 2013; two presentations 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 (CCW), 19 UNTS 1823 (1990), available at www.unog.ch/\_80256e e600585943.nsf/(httpPages)/a038dea1da906f9dc1257dd90042e261?OpenDocument&Expa ndSection=1#\_Section1 and www.unog.ch/80256EE600585943/(httpPages)/6CE049BE22E C75A2C1257C8D00513E26? and two forthcoming articles 'Autonomous weapons systems (AWS) and human rights law in the context of law enforcement' (*Human Rights Quarterly*, forthcoming 2016) and 'Autonomous weapons systems (AWS) and international law in the context of armed conflict'. I thank Thompson Chengeta and Petronell Kruger for their help with this contribution.

<sup>1</sup> See, e.g., E. Brynjolfsson and A. McAfee, *The Second Machine Age* (New York: Norton, 2014); N. Bostrom, *Superintellegence* (Oxford University Press, 2014). On the robotic weapons revolution, see P. Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century* (London: Penguin, 2009).



INTRODUCTION

over the millennia to determine when one human may use force against another, in peace and in war, and have assigned responsibility for violations of these norms. Perhaps the most dramatic manifestation of the rise of computer power is to be found in the fact that we are on the brink of an era when decisions on the use of force against human beings – in the context of armed conflict as well as during law enforcement, lethal and non-lethal – could soon be taken by robots.

Unmanned or human-replacing weapons systems first took the form of armed drones and other remote-controlled devices, which allowed human beings to be physically absent from the battlefield. Decisions to release force, however, were still taken by human operatives, albeit from a distance. The increased autonomy in weapons release now points to an era where humans will be able to be not only physically absent from the battlefield but also psychologically absent, in the sense that computers will determine when and against whom force is released. The depersonalization of the use of force brought about by remote-controlled systems is thus taken to a next level through the introduction of the autonomous release of force.

Drones raise a host of questions, but what is asked in this chapter is how the international community should respond to the new development outlined above: autonomous weapons systems (AWS), which could soon have the power to inflict serious physical injury, or even death, on human beings? AWS may be defined as robotic weapons that, once activated, can select and engage targets without further human intervention.<sup>2</sup> They have sensors that provide them with a degree of situational awareness, computers that process the information, and effectors (weapons) that implement the 'decisions' taken by the computers.

It should be made clear that what is at stake here are decisions over critical functions – determinations about the release of force – and not decisions over other functions such as navigation and takeoff and landing. Moreover, the force at stake in this discussion is force that is used against human beings – not force used against objects such as incoming

<sup>&</sup>lt;sup>2</sup> US Department of Defense (DoD) Directive 3000.09, 'Autonomy in weapon systems', 21 November 2012, Glossary part II, available at www.dtic.mil/whs/directives/corres/pdf/ 300009p.pdf; Human Rights Watch, Losing Humanity: The Case against Killer Robots (2012), 2, available at www.hrw.org/sites/default/files/reports/arms1112ForUpload\_0\_0.pdf; see also United Kingdom Ministry of Defence (MoD), 'The UK approach to unmanned aircraft systems', Joint Doctrine Note 2/11, 30 March 2011, paras. 202–3, available at www.gov.uk/government/publications/jdn-2-11-the-uk-approach-to-unmanned-aircraft-systems.



#### AUTONOMOUS WEAPONS SYSTEMS

munitions or other robots, which does not raise the same considerations of intrusion on people's bodies and lives.<sup>3</sup> At the same time, the issue addressed is not confined to the use of lethal force; killing and injuring people raise largely the same questions of infringement of bodily security and are both under consideration here, although lethal force clearly constitutes the extreme case and, as such, receives the bulk of the attention.<sup>4</sup>

Increased levels of autonomy in the use of AWS in force delivery occur especially in the military context. A low level of machine autonomy, which is clearly subordinate to human autonomy, may in some cases be involved. An example would be the computer programs that suggest targets and angles of attack to drone operators. On the other side of the spectrum, there are weapons whereby machines essentially take the targeting decisions out of human hands. One example would be the longrange anti-ship missile, a precision-guided anti-ship standoff missile with autonomous targeting capabilities that can detect and destroy specific targets within a group of numerous ships at sea. Full machine autonomy has not yet been used against human targets, but the point has been reached where that possibility has become very real.

<sup>3</sup> See Jewish Virtual Library, available at www.jewishvirtuallibrary.org/jsource/Peace/Iron Dome.html; US Military, available at http://usmilitary.about.com/library/milinfo/navy facts/blphalanx.htm.

- <sup>4</sup> Should the issue at stake in discussions about AWS be the use of lethal force by such weapons or any use of force against human beings? Since the debate on these new weapons started out as one on its military implications, it took a certain direction. What was discussed was the infliction of death in the context of armed conflict, where the use of force is regulated with reference to international humanitarian law and the use of deadly force against legitimate targets is the norm. Terminology such as 'lethal autonomous weapons' is used. See CCW Meeting on Lethal Autonomous Weapons Systems, available at http://bit.ly/1jSlCro. I used the term 'lethal autonomous robotics' in my report in 2013. Heyns, Doc. A/HRC/23/47. This terminology may suggest an approach in terms of which this was seen as an international humanitarian law issue, and the appropriate international for afor the discussion of AWS were considered to be disarmament bodies. With time, however, there was a realization that the underlying issue was the use of autonomous force against human beings in general, also during policing operations, when the use of graduated force is required and deadly force is the exception. The same issues of bodily integrity and human dignity arise, and, in many cases, it is impossible to foresee in advance whether force will be lethal or not. As such, the more inclusive term 'autonomous weapons systems' appears to be more appropriate.
- 5 See P. Scharre, 'Autonomy: killer robots and human control in the use of force', part 2 (2014), available at http://justsecurity.org/12712/autonomy-killer-robots-human-control-force-part-ii.
- $^{\rm 6}\,$  See Lockheed Martin, available at www.lockheedmartin.com/us/products/LRASM.html.



6

Cambridge University Press 978-1-107-15356-1 — Autonomous Weapons Systems Edited by Nehal Bhuta, Susanne Beck, Robin Geiβ, Hin-Yan Liu, Claus Kreβ More Information

Most of the unmanned systems that are becoming available to those who engage in ordinary law enforcement are remote controlled.7 However, some level of autonomy in force release is also becoming available for law enforcement purposes. For example, a fixed automatic tear gas system that can be fitted to police barriers during demonstrations is available. It releases doses of tear gas if a perpetrator ignores the warning and penetrates further into a restricted area. Some automation is also present in so-called automated rifle scoping, where a computer

INTRODUCTION

decides when to release fire against a human-selected target. The computer increases the first shot success probability – and, thus, diminishes the chances of bystanders being hit - by releasing fire only when a trajectory has been found that compensates for the effect of gravity, wind and so on.9

AWS, whether used in armed conflict or law enforcement, are weapon platforms, and any weapon can in principle be fitted onto an AWS. Therefore, the important distinguishing feature between different kinds of AWS is not the weapons they use but, rather, how they take their decisions - their levels of autonomy. While some AWS operate at low levels of autonomy, under close human control, it is also clear that some AWS will be able to operate at high levels of independence. It has become customary to refer to those systems in which there is no meaningful human control over force release as 'fully autonomous' weapons systems.

There are a number of reasons why AWS are being developed. The primary rationale for the development of unmanned systems in general (remote-controlled weapons and AWS) is their ability to protect personnel who are kept out of harm's way. However, what motivates the additional move from remote control to autonomous weapons release? The main argument is that AWS may be quicker at engaging intended targets because they can process information faster.<sup>10</sup>

scope-of-smart-weapons.

<sup>&</sup>lt;sup>7</sup> E.g., police UAV drones, remote aerial platform tactical reconnaissance, available at www .policeuavdrones.com/.

8 See Security Research Map, available at www.securityresearchmap.de/index.php?lan

g=en&contentpos=4046. See also Roto Concept, available at www.rotoconcept.com.

9 See Motherboard, available at http://motherboard.vice.com/blog/long-shot-inside-the-

T. K. Adams, 'Future warfare and the decline of human decision making', *Parameters*: United States Army War College Quarterly, 31(4) (2001), 57-8; G. E. Marchant et al., 'International governance of autonomous military robots', Columbia Science and Technology Law Review, 12 (2011), 280; Heyns, Doc. A/HRC/23/47, 8.



## AUTONOMOUS WEAPONS SYSTEMS

Furthermore, it is also sometimes argued that the superior processing powers of AWS can prevent the wrong targets from being hit. <sup>11</sup> To the extent that this argument is correct, the increased depersonalization in the deployment of force brought about by AWS may thus lead to greater personalization in targeting outcomes and saving lives or preventing unwarranted injuries. For example, in an armed conflict, robots that are programmed to return fire may be able to engage in more incisive exploration as to whether a perceived threat is real before it uses force than its human counterpart can. Humans in such a situation may be inclined to shoot earlier out of fear and, in the process, kill civilians who are not engaged in hostilities. <sup>12</sup>

Increasingly autonomous weapons have been developed largely in the military context, but some of the same arguments in favour of AWS may also be used in the law enforcement context. Automation of force can arguably allow greater speed and accuracy in targeting or preventing the excessive use of force. This potential could be used, for example, in a hostage situation where all of the hostage takers need to be hit at the same time to protect the lives of the hostages. In both the military and policing contexts, robots can do the dull, dangerous and dirty work.

However, there are serious questions to be asked on the use of AWS. The ethical and legal considerations applicable to the use of force against human beings today are largely expressed through the use of human rights language. International law has formulated a number of explicit rules that determine when such force may be used, and there is broad ethical support for this approach. Human rights law stipulates as a general rule that one person may use force against another only where the interest protected outweighs the harm done (that is, the force used is proportionate) and there is no other way to prevent such harm (the force used must be necessary) and it may only be used against an imminent attack.

When lethal force is used by law enforcement officials – that is, the harm done is that someone is killed – the proportionality requirement can only be fulfilled if the interest that is protected is the life of another

See, e.g., B. J. Strawser, Killing by Remote Control: The Ethics of an Unmanned Military (Oxford University Press, 2013), 17. See also M. Horowitz and P. Scharre, 'Do killer robots save lives?' available at www.politico.com/magazine/story/2014/11/killer-robots-save-lives-113010.html.

R. C. Arkin, 'Lethal autonomous weapons systems and the plight of the non-combatant' (2014), 3, available at www.unog.ch/80256EDD006B8954/%28httpAssets%29/54B1B7 A616EA1D10C1257CCC00478A59/\$file/Article\_Arkin\_LAWS.pdf.



.

#### INTRODUCTION

person. This has been called the 'protect life' principle: as a general standard, a life may only be taken if it is absolutely necessary to protect another life. The 'protect life' principle is the guiding star whenever lethal force is used. However, the exceptional circumstances that prevail during armed conflict – such as the difficulty of exercising control over the use of force over a long distance and the fog of war – make it, on a temporary basis, an impossible standard to enforce. During armed conflict, human rights law remains valid, but it is interpreted with reference to the rules of international humanitarian law (IHL). 14

In essence, the rules of IHL are those of distinction (only legitimate targets may be attacked); proportionality (any incidental or collateral damage inflicted on civilians who are not directly participating in hostilities must not be excessive in relation to the military advantages obtained); and precaution (feasible precautions must be taken to protect civilians). These are the explicit rules of international law. However, it could also be argued that it is an implicit assumption of international law and ethical codes that humans will be the ones taking the decision whether to use force, during law enforcement and in armed conflict. Since the use of force throughout history has been personal, there has never been a need to make this assumption explicit. The advent of AWS makes addressing this issue now a priority. I will briefly address the three main questions raised by AWS.

## 'Can they do it?'

Can AWS, as a practical matter, meet the explicit rules regarding the use of force, as set out earlier, in the context of law enforcement or armed conflict? The right to life may potentially be infringed by AWS in

Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, ICJ Reports (1996) 226, para. 25, available at www.refworld.org/docid/4b2913d62.html.

For a discussion, see, e.g., S. Oeter, 'Methods and means of combat: I General rules' in D. Fleck (ed.), *The Handbook of International Humanitarian Law*, 2nd edn (Oxford University Press, 2008), 130, 134; J-M. Henckaerts and L. Doswald-Beck, *Customary International Humanitarian Law*, vol. 1: *Rules*, Rules 7, 14–24 (Cambridge University Press, 2005), 25–9, 46–76.

See Principle 9 of the Basic Principles on the use of Force and Firearms by Law Enforcement Officials, available at www.unrol.org/files/BASICP~3.PDF, adopted by the Eighth United Nations Congress on the Prevention of Crime and the Treatment of Offenders, Havana, Cuba, 27 August – 7 September 1990. It is a minimum standard, in that not all uses of deadly force in order to save life are justified. It may, for example, not be acceptable to kill innocent bystanders to save someone on the proverbial runaway trolley.



AUTONOMOUS WEAPONS SYSTEMS

Q

a number of ways if they are used in these contexts.<sup>16</sup> The most often cited instance is when the deployment of force through AWS results in the direct use of force against those who are not considered to be legitimate targets under the law.<sup>17</sup> The wrong people may be hit because computers may not be able to identify the correct targets. Clearly, autonomous technology can assist in simple cases, but more complicated decisions may specifically require human judgment.

In law enforcement situations, proper targeting may require an understanding of human intentions. The determination whether there is an imminent attack that warrants the use of force may simply not be made properly by computers, now or in the future. In the case of armed conflict, machines, for example, may not always be able to distinguish those who are wounded or in the process of surrendering from those who may legitimately be targeted. They also may not be able to differentiate between civilians who are directly participating in hostilities and those who are not.

Even if the force used is not misdirected, the level of the force used by AWS may still be excessive. In law enforcement, any force used must be the minimum required by the circumstances. During armed conflict, excessive force may manifest itself in unacceptably high levels of collateral or incidental casualties. It is difficult to imagine that machines will ever be able to take such decisions in a reliable way. Again, the concern is that robots may not be able to make the essentially qualitative, often value-based, decisions that are required to ensure that such force is not excessive.

According to Article 6(1) of the International Covenant on Civil and Political Rights 1966, 999 UNTS 171, '[e] very human being has the inherent right to life. This right shall be protected by law. No one shall be arbitrarily deprived of his life.'

<sup>17</sup> In terms of Article 48 of Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts to the Geneva Conventions (Additional Protocol I) 1977, 1125 UNTS 3, to 'ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives'. At no time shall parties make civilians the object of attack.

Article 51(5)(b) of Additional Protocol I to the Geneva Conventions: 'Among others, the following types of attacks are to be considered as indiscriminate: (b) an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.' Excessive collateral damage may also occur where no force should have been used at all, because excessive incidental casualties are unavoidable.



10

#### INTRODUCTION

# 'Should they do it?'

However, even if we were to assume that the answer to the first question is affirmative – if it is accepted that AWS can engage in reasonably accurate targeting – a further question presents itself: is it right for machines to have the power of life and death over humans or the ability to inflict serious injury? This question brings us back to the argument that there may be an implicit requirement in terms of international law and ethical codes that only human beings may take the decision to use force against other humans. The implication of this approach bears emphasis. If there is such a requirement, then even if the correct target is hit and the force used is not excessive – and, in that sense, the explicit requirements of international law are met in a formal way – it will remain inherently wrong for a machine to make the determination that such force be used against a human being.

Seen from this perspective, it could be an inherently arbitrary deprivation of the right to life if the decision to use deadly force is delegated to machines. Human life, it has been argued, can only be taken as part of a process that is potentially deliberative and involving human decision making. While the infliction of deadly force, especially during armed conflict, is often not deliberative in practice, AWS, which implies a high degree of autonomy, decisively rules out that possibility.

I would like to advance a further consideration to support the above contention, namely the implications of fully autonomous AWS for the right to dignity. To allow such machines to determine whether force is to be deployed against a human being may be tantamount to treating that particular individual not as a human being but, rather, as an object eligible for mechanized targeting.<sup>20</sup> It should be recognized that the exact contents and interpretation of the right to dignity is contested, and some call it a conversation stopper.<sup>21</sup> What cannot be contested, however, is that the concept of dignity has played a central role and

P. M. Asaro, 'On banning autonomous weapon systems: human rights, automation and the dehumanisation of lethal decision making', *International Review of the Red Cross*, 94 (886) (2012), 2, 8–17. See also P. M. Asaro, 'Robots and responsibility from a legal perspective', available at www.peterasaro.org/writing/ASARO%20Legal%20Perspective .ndf.

<sup>.</sup>pdf.

R. Sparrow, 'Robotic weapons and the future of war' in J. Wolfendale and P. Tripodi (eds.), New Wars and New Soldiers: Military Ethics in the Contemporary World (Farnham: Ashgate, 2012), 11; A. M. Johnson, 'The morality of autonomous robots', Journal of Military Ethics, 134 (2013), 134; Heyns, Doc. A/HRC/23/47, 18, para. 95.

D. Birnbacher, 'Ambiguities in the concept of Menschenwürde' in K. Bayertz (ed.), Sanctity of Life and Human Dignity (New York: Springer, 1996), 107.



## AUTONOMOUS WEAPONS SYSTEMS

served as a driving force in the development of both human rights law<sup>22</sup> and IHL.<sup>23</sup> It can be expected that it will and should continue to play a role in both of these branches of law when new challenges are confronted.

Dignity, at least in the Kantian tradition, advances the idea of the infinite or incommensurable value of each person. It has been argued that to have the decision whether you live or die – or be maimed – taken by machines is the ultimate indignity. Robots cannot be programmed to respond in an appropriate way to the infinite number of possible scenarios that real life – and real people – offers. Death by algorithm means that people are treated simply as targets and not as complete and unique human beings, who may, by virtue of this status, deserve to meet a different fate.

A machine, which is bloodless and without morality or mortality, cannot do justice to the gravity of the decision whether to use force in a particular case, even if it may be more accurate than humans.<sup>27</sup> This decision is so far-reaching that each instance calling for its use requires that a human being should decide afresh whether to cross that threshold if it is not to become a mechanical – and inhuman – process. Moreover, the dignity of those in whose name fully autonomous AWS are used may also be implicated. When such AWS dispense force on their behalf, they may not be able to act as moral agents who take their own

- Dignity has been called the 'mother' of human rights. See B. Schlink, 'The concept of human dignity: current usages, future discourses' in C. McCrudden (ed.), *Understanding Human Dignity* (Oxford University Press, 2013), 632. See also P. Carozza, 'Human dignity and judicial interpretation of human rights: a reply', *European Journal of International Law* 19 (5) (2008), 931–44, available at http://ejil.oxfordjournals.org/content/19/5/931.full; Nils Petersen, 'Human dignity, international protection', available at http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e809 ?print.
- See, e.g., the view of the International Committee of the Red Cross (ICRC) as expressed in www.icrc.org/eng/resources/documents/article/other/ihl-human-rights-article-011207.htm. See also Rule 90 of the ICRC's Customary Law Study in Henckaerts and Doswald-Beck, Customary International Humanitarian Law, vol. 1; Geneva Conventions, Common Article 3(1)(c); Additional Protocols I, Article 75(2) and preamble and Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts 1978, 1125 UNTS 609, Article 4(2).
- R. J. Scott, 'Dignité/dignidade: organising against threats to dignity in societies after slavery' in McCrudden, Understanding Human Dignity, 69.
- Major General R. H. Latiff and P. McCloskey, 'With drone warfare, America approaches the robo-rubicon', Wall Street Journal (14 March 2013).
- See Human Rights Watch, 'Shaking the foundations: the human rights implications of killer robots' (2014), 2, available at www.hrw.org/sites/default/files/reports/arms0514\_ForUpload\_0.pdf.
- <sup>27</sup> See Asaro, 'On banning autonomous weapon systems', 695; see also 689, 694–700.



12

#### INTRODUCTION

decisions; instead, they abdicate their moral responsibility to bloodless entities. They are unable to assume, and exercise, responsibility.

# Issues of accountability

The modern concept of human rights entails that certain core values are protected, and, if they are violated, there is accountability. Accountability is part of the protection of a particular right. A lack of accountability for a violation of the right to life, for example, is in itself a violation of that right. Accountability can take many forms, including criminal prosecution, civil damages, disciplinary steps or the offering of redress or compensation. In addition to individual responsibility, institutions, such as states or corporations, may be held accountable. To the extent that this approach is to be followed in the case of AWS, the depersonalized use of force will result in depersonalized forms of responsibility.

Accountability is traditionally premised on control. For example, under criminal responsibility, one cannot be held responsible for that which is outside your control. To the extent that AWS allow for, and are under, human control, humans will remain responsible. However, to the extent that AWS are outside human control, it appears that there may be an accountability vacuum. There is clearly no point in putting a robot in jail.

Even where humans as a collective do exercise significant control over AWS, there may be uncertainty about whether each – or any – individual should be held to account in a specific case. Some form of human control over AWS may be exercised on different levels in the wider loop – for example, through computer programming, through the decision that the military will use such weapons, through the commander who orders subordinates to do so, and so on. It is not clear how responsibility should be assigned in such cases and to what extent it may be appropriate to hold

Human Rights Committee (HRC), General Comment no. 31: 'The nature of the general legal obligation imposed on states parties to the covenant', UN Doc. CCPR/C/21/Rev.1/Add.13 (2004) adopted on 29 March 2004, paras. 16 and 18; see also HRC, General Comment no. 6, UN Doc. HRI/GEN/1/Rev.1 at 6 (1994); 'Basic principles and guidelines on the right to a remedy and reparation for victims of gross violations of international human rights law and serious violations of international humanitarian law', Resolution 60/147 adopted by the General Assembly on 16 December 2005, A/60/509/Add.1 adopted and proclaimed by UN General Assembly Resolution 60/147 of 16 December 2005, para. 4; ECtHR, McCann and others v. The United Kingdom, Appl. no. 18984/91, judgment of 27 September 1995, para. 169.