

# Autonomous weapons in armed conflict and the right to a dignified life: an African perspective

Christof Heyns\*

Institute for International and Comparative Law in Africa, University of Pretoria, Pretoria, South Africa

\*Correspondence author: Christof Heyns

Institute for International and Comparative Law in Africa, University of Pretoria, Lynnwood Road, Pretoria 0002, South Africa. christof.heyns@up.ac.za

## Abstract

Autonomous weapons are weapons that, once activated, can without further human intervention select and engage targets. This raises the possibility that computers will determine whether people will live or die. The possible use of autonomous weapons against humans in armed conflict clearly has potential right to life implications. This contribution argues that the right to dignity angle must also be brought into play. The first concern raised by autonomous weapons is 'can they do it?': Can autonomous targeting conform with the requirements of international humanitarian law, in particular the rules of distinction and proportionality? If machines cannot do proper targeting, such use of force will be 'arbitrary' and thus violate the right to life. Moreover, the right to life requires accountability, but it is not clear who is to be held responsible when robots get it wrong. Secondly: 'Should they do it?' Should robots have the power of life and death over humans? This may violate the rights to life as well as the right to dignity. The question whether there is 'meaningful human control' over the release of force is emerging as a helpful tool to distinguish between acceptable and unacceptable autonomous targeting, and I argue that it also makes sense from a human rights perspective. The question that will haunt the debate in the future is: What if technology develops to the point where it is clear that fully autonomous weapons surpass human targeting, and can potentially save many lives? Would human rights considerations in such a case not militate for the use of autonomous weapons, instead of against it? I argue that the rights to life and dignity demand that even under such circumstances, full autonomy in force delivery should not be allowed. The article emphasises the importance placed on the concept of a 'dignified life' in the African human rights system.

**Keywords:** International humanitarian law; autonomous weapons; international law; human rights

## 1. Introduction

Traditionally, weapons users have been in direct physical control of their weapons. Over the years, revolutions in military affairs have produced weapons with increased range and fire-power, but, by and large, this did not change the fact that the person

activating or launching the weapon took the decision when and against whom force would be used while being present on the spot from where force was projected.<sup>1</sup>

The advent of unmanned or human-replacing systems has meant that the person launching the weapon no longer needs to be physically present at the place and time when it is released. The first generation of these unmanned systems are remote-controlled weapons systems. The best-known iteration of this technology, armed drones, allow their users to be half-way around the world in a control room when pressing a button that releases a missile from a weapon platform hovering above the target.<sup>2</sup>

The emphasis of this article is on the next generation of unmanned systems in armed conflict to which developers have now turned their attention. So-called 'autonomous weapons' (sometimes simply called robots or machines) would allow for the release of force from unmanned systems that are not directly or remotely controlled by humans.<sup>3</sup> Instead, once a human has activated an autonomous weapon system, on-board computers will make the determination, independent from direct human intervention, against whom it should be directed and when to release force. In other words, the machine itself will exercise these 'critical functions'. Humans remain in the wider decision-making loop, but a computer plays a significant role in the selection of the target and release of force.<sup>4</sup>

The emergence of autonomous weapons will make it possible for humans not only to be physically absent from the point of release of force, as with armed drones, but to be 'psychologically' absent as well, to the extent that they do not take the on-the-spot decision to direct and open fire. The actual release of force takes place elsewhere and some time after the system was activated. There is thus a process of increasing depersonalisation in the use of force through unmanned systems – already present in the case of remote-controlled weapons, but taken to the next level with autonomous weapons. The force projected by such weapons could of course be directed against objects. However, our concern is confined to human rights and in particular, the right to life and the right to dignity, and as a result I focus on the situation where robots use force against human beings.

Autonomous weapons are currently in various stages of development. Although the use of fully autonomous weapons is not yet officially sanctioned as such by any government, there are already borderline cases, and unless there is a conscious collective decision to the contrary, there is every reason to believe that it will soon be in use. As human beings, we have grappled since time immemorial with making sense of death at the hands of other humans or as a result of nature. We are now confronted

---

<sup>1</sup> On revolutions in military affairs and the rise of unmanned systems, see P Singer *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (2009).

<sup>2</sup> On the legal framework related to drones, see C Heyns and others 'The International Law Framework Regulating the Use of Armed Drones' (2016) 65 *International and Comparative Law Quarterly* 79–827.

<sup>3</sup> Autonomous weapons are platforms, to which any weapons may be fitted. In armed conflict the weapons used are as a rule lethal – hence the term Lethal Autonomous Weapons Systems is sometimes employed in this context. It is also possible that less-lethal weapons may be fitted to an autonomous weapon platform for use in law enforcement. The term 'autonomous weapons' (thus dropping the word 'lethal') could be an appropriate umbrella term that serves to cover its use during armed conflict as well as law enforcement. Using this more inclusive term may help to emphasise that some of the same issues arise in both situations, and that these weapons should not only be dealt with as a disarmament matter and in disarmament fora.

<sup>4</sup> For a discussion of the definition of autonomous weapons, see M Schmitt 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) 1 *Harvard National Security Journal Features* 3–7.

with a new case: death by algorithm. This challenges many of the assumptions engrained in the ethical, legal, religious and other normative systems of the world.

The ‘autonomy’ of robots is not comparable to the autonomy of human beings, which is often seen as the basis of the ability of humans to act as free moral agents. However, robots with a high level of autonomy can perform functions that those who programme or deploy them cannot foresee (and in that sense, autonomous weapons have something that might resemble ‘free will’). ‘Automatic’ systems respond in predictable ways to their environment. ‘Autonomous’ systems, however, may sometimes respond in unpredictable ways, and are, to the extent that this happens, outside human control. Unpredictability may occur because machine learning takes place (they alter their behaviour based on their experience). Moreover, especially in the context of something as chaotic as war, not all scenarios can be foreseen and be programmed into the computer of the machine beforehand. As such, machine autonomy, beyond a certain point, can potentially undermine or limit human autonomy and control over the world.

While earlier revolutions in military affairs gave the warrior control over ever-more-powerful weapons, autonomous weapons have the potential to bring about a change in the identity of the decision-maker. The weapon may now become the warrior.<sup>5</sup>

According to philosopher Nick Bostrom, the appropriate response of human beings to the advent of artificial intelligence in all walks of life is the central question of the day.<sup>6</sup> Computers already exercise far-reaching power in our lives. Algorithms play a key role, for example, in transactions on the stock exchange and whether banks grant credit or close accounts. It also deeply influences the exercise of public power, in decisions on matters such as taxation and the granting of social benefits and many others. Against a context where digitalisation increases the potential of state control over society, the question arises how far the process of transferring power from humans to machines should go.

Autonomous weapons represent this question perhaps in its starkest form. In the context of a more general concern about the role of computers in the world of the future, the significance of the possibility that they may soon be able to forcibly end human life is clear. If machines hold the key to the use of deadly violence against humans, the question can be asked why they should not take almost any other decision – ranging from the most mundane to those that may determine whether human life is sustained (e.g. in the context of euthanasia) or deeply affect our lives in other ways (e.g. by taking over the functions of courts of governments, including in the context of the distribution of resources). The way we respond to autonomous weapons is a pivotal test case for the role of science in the future. The stakes cannot be higher – they are literally life and death – and how we deal with autonomous weapons will be the tone for how we deal with computers in general.<sup>7</sup>

From what normative perspective can this issue be assessed? An important perspective would clearly be the one of human rights. Human rights are widely accepted

---

<sup>5</sup> See the discussion by R Crotoof ‘Autonomous Weapon Systems and the Limits of Analogy’ (2016) <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2820727&download=yes](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820727&download=yes)> accessed 21 March 2017.

<sup>6</sup> This is the central argument advanced in N Bostrom *Superintelligence: Paths, Dangers, Strategies* (2014).

<sup>7</sup> C Heyns ‘Autonomous weapons systems: Living a dignified life and dying a dignified death’ in N Bhuta, S Beck, R Geiß, H Liu & C Kreß (eds) *Autonomous Weapons Systems: Law, Ethics, Policy* (2016) 13.

as the dominant normative – ethical and legal – framework of the international community today. Louis Henkin has famously called human rights ‘the idea of our time’.<sup>8</sup> The rise of autonomous weapons thus presents the opportunity to ask what the implications of the dominant normative framework of our time are for the central question of the day.

No doubt the first question from a human rights perspective should be to what extent do autonomous weapons affect the right to life, often described as the ‘supreme right’.<sup>9</sup> They may, for example, not be able to make the value judgements required to ensure targeting outcome that are consistent with the rules of international humanitarian law (IHL), and thus violate the rights to life of those who are directly targeted or indirectly affected as incidental casualties. Less often discussed, however, is the extent to which autonomous weapons may also affect the right to dignity, a less clearly defined right but one that has played a foundational role in the development of both human rights law and IHL.<sup>10</sup> Is it not an affront to human dignity if robots have the power of life and death over humans? It is through the prism of these two pillars of the human rights project – the right to life and the right to dignity – that I will look at autonomous weapons.

The potential advantages of some level of autonomy in targeting should be clearly understood. Unmanned systems in general offer the advantage of the protection of one’s own forces – a significant advantage, also from the point of view of the right to life. Autonomous systems offer clear additional potential military advantages to those who deploy them, such as increased reaction speed in decision-making and ‘eyes’ on the target, potentially increasing accuracy. Machines may in some cases avoid mistakes that humans make and even atrocities that they commit on the battlefield because of emotions such as fear, fatigue, or revenge. It should be recognised that human compliance with IHL is appallingly low, and if technology can improve that, the gains will be significant.<sup>11</sup>

If using autonomy in weapons can lead to better targeting, as proponents claim, autonomous weapons can save the lives of those who may not be targeted, such as civilians not engaged in the conflict, and reduce collateral damage. It is on that basis that roboticist Arkin and others has defended the development and use, under certain circumstances, of autonomous weapons.<sup>12</sup> Clearly, such potential capabilities are important from a military as well as human rights perspective. This is not a simple matter where a blunt instrument like a blanket ban on all forms of autonomy in force delivery is likely to yield satisfactory results. We should rather ask whether it is

---

<sup>8</sup> L Henkin *Age of Rights* (1990) ix.

<sup>9</sup> UN Human Rights Committee CCPR General Comment No.6: The Right to Life (Article 6) (30 April 1982) para 1. See also Human Rights Watch *Shaking the Foundations: The Human Rights Implications of Killer Robots* (2014) 6 [https://www.hrw.org/sites/default/files/reports/arms0514\\_ForUpload\\_0.pdf](https://www.hrw.org/sites/default/files/reports/arms0514_ForUpload_0.pdf) accessed 21 March 2017.

<sup>10</sup> See Common Article 3 to the Geneva Conventions of 1949; Article 75(2) of Additional Protocol I and Article 4(2) Additional Protocol II to the Geneva Conventions of 1949; Article 8(2)(c)(i) and (ii) of Statute of the International Criminal Court; Article 4(e) of the International Criminal Tribunal for Rwanda; Article 3(e) of the Statute of Special Court for Sierra Leone; Rule 90 of ICRC Customary Law Study; *Prosecutor v Aleksovski*, Case No IT-95-14/1-T, Judgement, 25 June 1999, para 54; *Prosecutor v Aleksovski*, Case IT-95-14/1-A, Judgement (24 March 2000) 26; *Prosecutor v Zejnil Delalic, Zdravko Mucic (a/k/a/ 'Pavo'), Hazim Delic, Esad Landzo (a/k/a 'Zenga')* Case No. IT-96-21-T paras 512–44.

<sup>11</sup> See, in general, R Arkin ‘Lethal Autonomous Systems and the Plight of the Non-combatant’ (2014) 1 *Ethics and Armed Forces* 3.

<sup>12</sup> Arkin (note 11) 9.

possible to distinguish acceptable from unacceptable uses of autonomy. To what extent is it possible to ensure that the benefits offered by autonomous weapons are used, but its disadvantages and in particular the possible infringements of human rights alluded to earlier are avoided?

At an early stage of the debate, I presented some of the questions raised by autonomous weapons as Special Rapporteur on extrajudicial, summary or arbitrary executions in a report to the United Nations (UN) Human Rights Council in 2013.<sup>13</sup> I argued that states should impose national moratoria on the development and use of autonomous weapons until such time as an internationally acceptable way of dealing with increasing autonomy in targeting has been found. The global engagement with this issue has now advanced to the point where it may be easier to come to a firm view. This issue has been taken up in a number of international fora and has been subjected to thorough multi-disciplinary consideration.<sup>14</sup> Eminent scientists have expressed themselves on the topic.<sup>15</sup> A thorough academic discussion has also taken place.<sup>16</sup> Significantly, the state parties to the Convention on Conventional Weapons have been engaged with the issue on an annual basis since 2014 and in 2016 decided to establish a Group of Governmental Experts who will meet during 2017 to further consider the way forward. Some have argued that this may lead to a prohibition on some forms of autonomous targeting.<sup>17</sup>

There is a growing view that the dividing line between acceptable and not acceptable machine autonomy is determined by the question whether humans retain ‘meaningful human control’ over force delivery. Where humans are the real decision makers to the extent that accountability can be assigned should something go wrong, autonomous weapons are acceptable, but where there is no meaningful human control – where there is in other words ‘full autonomy’ on the part of the machine – they should be banned. Some call such robotic weapons without meaningful human control ‘killer robots’.<sup>18</sup>

---

<sup>13</sup> See A/HRC/23/47, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, C Heyns (9 April 2013).

<sup>14</sup> The United Nations Disarmament Committee Expert Meeting on AWS (2014) <<https://www.un.org/disarmament/geneva/ccw/2014-meeting-of-experts-on-laws/>>; (2015) <[http://www.unog.ch/80256EE600585943/\(httpPages\)/6CE049BE22EC75A2C1257C8D00513E26](http://www.unog.ch/80256EE600585943/(httpPages)/6CE049BE22EC75A2C1257C8D00513E26)>; (2016) <[http://www.unog.ch/80256EE600585943/\(httpPages\)/37D51189AC4FB6E1C1257F4D004CAF2B?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/37D51189AC4FB6E1C1257F4D004CAF2B?OpenDocument)> accessed 21 March 2017.

<sup>15</sup> See ‘Autonomous Weapons: An Open Letter from AI & Robotics Researchers’ (2015) <<http://futureoflife.org/open-letter-autonomous-weapons/>> accessed 21 March 2017.

<sup>16</sup> See K Anderson & M Waxman ‘Law and Ethics for Autonomous Weapon Systems: Why a Ban Won’t Work and How the Laws of War Can’ American University Washington College of Law Research Paper No. 2013-11; K Anderson & M Waxman ‘Law and Ethics for Robot Soldiers’ American University Washington College of Law Research Paper No. 2012-32; R C Arkin ‘Governing Lethal Behaviour in Autonomous Robots’ (2009) *International Committee of the Red Cross Press*; P Asaro ‘*Jus Nascendi*, Robotic Weapons and the Martens Clause’ (2015); J Lewis ‘The Case for Regulating Fully Autonomous Weapons’ (2015) *Yale Law Journal* 1309; W C Marra & S K McNeil ‘Understanding ‘The Loop’: Regulating the Next Generation of War Machines’ (2013) 36 *Harvard Journal of Law and Public Policy* 1139; N Melzer ‘Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare’ (2013) European Parliament Directorate-General for External Policies, EXPO/B/DROI/2012/12; N E Sharkey ‘The Evitability of Autonomous Robot Warfare’ (2012) *International Review of the Red Cross*; R Sparrow ‘Killer Robots’ (2007) 24 *Journal of Applied Philosophy* 62; J S Thurnher ‘No One at the Controls: Legal Implications of Fully Autonomous Targeting’ (2012) 67 *Joint Force Quarterly* 77; M Wagner ‘Taking Humans Out of the Loop: Implications for International Humanitarian Law’ (2011) 21 *Journal of Law Information and Science* 155.

<sup>17</sup> For further reading on this argument, see <[http://www.unog.ch/80256EE600585943/\(httpPages\)/9F975E1E06869679C1257F50004F7E8C?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/9F975E1E06869679C1257F50004F7E8C?OpenDocument)> accessed 21 March 2017.

<sup>18</sup> See ‘Campaign to Stop Killer Robots’ <<http://www.stopkillerrobots.org/>> accessed 21 March 2017.

In order to ascertain more clearly to what extent the notion of ‘meaningful human control’ addresses the concerns relating to the right to life and the right to dignity, it is useful to articulate the two primary questions raised by such weapons in the context of armed conflict. The first question is whether they can do or enable proper targeting? – targeting of which the outcome compares favourably with that of targeting done by humans in terms of compliance with the rules of IHL.

For some, the inquiry stops here. To the extent that machines cannot enable proper targeting they should not be used; to the extent that they can do such targeting, there should not be any special impediments. However, others – and I am one of them – insist that a second question also has to be asked. Even if they can do proper targeting, should machines hold the power of life and death over humans? If the answer to the second question is negative, autonomous weapons should not be used even if they can do what would otherwise be considered proper targeting. I will argue that machines with meaningful human control do not violate the right to life and dignity, and there are no special considerations why they should be banned. The deployment of fully autonomous weapons, on the other hand, violate the rights to life and dignity, and should as a result not be allowed.

The reasons why fully autonomous weapons violate the rights to life as well as dignity are best brought to the fore by discussing two questions about deploying these sorts of weapons to target human beings – ‘can they do it’ and ‘should they do it’?

## **2. Can they do it? Autonomous weapons and proper targeting**

Can autonomous weapons do or enable proper targeting? That is, can they be deployed in such a way that the results of such targeting match or surpass the outcomes of targeting by humans in terms of meeting the rules IHL sets for humans? Of particular concern in addressing this question from the human rights perspective are those rules that protect some people on the battlefield from direct targeting (such as uninvolved civilians and the wounded) and from indirect targeting (through the prohibition of excessive incidental or collateral casualties)? What if autonomous weapons do worse than humans in striking the right target, or cause excessive incidental harm?<sup>19</sup>

It will pose too high a standard to require autonomous weapons to comply 100% with the rules of IHL – humans do not do this and in fact make mistakes in a disconcerting number of cases. Machines are also bound to fail in some cases. The real question is thus not how do machines compare with the ideal but rather how do they compare with human beings in the same situation.

It is useful to imagine that some kind of Turing test is done, and experts compare the outcomes of machine and human targeting to see whether they can assign it to the one or the other.<sup>20</sup> Insofar as machines consistently match or do better than humans it can be said that they engage in proper targeting.

---

<sup>19</sup> See Human Rights Watch, *Losing Humanity: The Case against Killer Robots* (2012) 32 <[https://www.hrw.org/sites/default/files/reports/arms1112\\_ForUpload.pdf](https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf)> accessed 21 March 2017.

<sup>20</sup> The Turing test is ‘a test for intelligence in a computer, requiring that a human being should be unable to distinguish the machine from another human being by using the replies to questions put to both’ <<http://www.turing.org.uk/scrapbook/test.html>> accessed 21 March 2017.

The right to life is typically framed as a right against the ‘arbitrary’ deprivation of life, and an ‘arbitrary’ deprivation of life can be seen as taking of life in a way that is in violation of international law.<sup>21</sup> The right to life has two components: the prohibition on the ‘arbitrary’ taking of life, and accountability where that occurs.<sup>22</sup>

The right to life continues to apply during armed conflict. As far as the conduct of hostilities is concerned, the question whether a deprivation of life is ‘arbitrary’ and thus unlawful under international law is determined with reference to the rules of IHL.<sup>23</sup> Targeting those who are protected from the intentional and direct use of force, as well as indiscriminate targeting or excessive collateral damage are considered arbitrary deprivations of life and thus constitute violations of the right to life. At least the serious violations of IHL must be investigated and if appropriate there must be prosecution.<sup>24</sup> To the extent that any of these rules of IHL are not followed when autonomous weapons deploy lethal force, the right to life is violated.

The IHL rules of distinction, proportionality and precaution in attack are of specific importance in this regard.<sup>25</sup> These rules must be complied with as a collective in each attack. The IHL rule of distinction seeks to minimise the direct impact of armed attacks on civilians and others who are not (or no longer) participating in hostilities, by prohibiting the targeting of such people.<sup>26</sup> Commanders are allowed to target categories of people based on status or conduct; they are not required to identify specific individuals who pose an immediate lethal threat as targets, as is the case with the use of deadly force under human rights law.<sup>27</sup> In situations where autonomous weapons cannot reliably distinguish between, for example, combatants or other fighters and civilians, their use would be unlawful.<sup>28</sup>

There are several factors that could possibly impede the ability of autonomous weapons to operate according to the rule of distinction, including the technological inadequacy of existing sensors;<sup>29</sup> a robot’s inability to understand context; and the difficulty of translating IHL language and definitions of civilian and combatant into

---

<sup>21</sup> See, eg Article 6 of the International Covenant on Civil and Political Rights of 1966.

<sup>22</sup> *Kaya v Turkey*, European Court of Human Rights, application No. 22729/93 (19 February 1998) paras 86-92. and *McCann and Others v the United Kingdom*, European Court of Human Rights, application No. 18984/91 (27 September 1995) para. 161.

<sup>23</sup> *Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion*, ICJ Reports (1996) 226, International Court of Justice (ICJ), 8 July 1996. Human rights law applies to all use of force, including during armed conflict. IHRL regulates the use of force during ‘normal’ peace time situations, while IHRL as well as IHL apply during the exceptional case of the conduct of hostilities during armed conflict, when human rights obligations, such as the right to life, are determined with reference to the provisions of IHL.

<sup>24</sup> See, eg Article 146 Geneva Convention IV.

<sup>25</sup> See M N Schmitt & J S Thurnher ‘Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict’ (2013) 4 *Harvard National Security Journal* 531, 246.

<sup>26</sup> In terms of Art 48 of Additional Protocol 1 to the Geneva Conventions of 1949, 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. The principle of distinction is a rule of customary international law. See *Legality of the Threat on the Use of Nuclear Weapons (Advisory opinion)*, 1996, ICJ Rep 226 para 79.

<sup>27</sup> See M Sassòli ‘Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified’ (2014) 90 *International Law Studies /Naval War College* 308, 315, 326, 333.

<sup>28</sup> Rule 1 of the Rules of Customary International Humanitarian Law requires that ‘parties to the conflict must at all times distinguish between civilians and combatants’. See also Article 8(2)(b)(xx) of the Rome Statute of the International Criminal Court of 1998.

<sup>29</sup> N Sharkey ‘Grounds for Discrimination: Autonomous Robot Weapons’ (2008) *RUSI Defence Systems* 88-9 <<https://www.scribd.com/document/96464854/23-Sharkey>> accessed 21 March 2017.

computer programming.<sup>30</sup> It may be difficult for robots to establish, for example, whether someone is a member of the armed forces or a civilian who participates directly in hostilities (and who may thus be targeted), as opposed to a law enforcement official or a hunter who carries a gun, or someone who is wounded and *hors de combat*, or is in the process of surrendering (and who may not be targeted).<sup>31</sup> Often, value judgements are involved, and even if the test is in principle objective, subjective factors will in many cases play a role in determining whether the decision-maker gets it right.<sup>32</sup> Humans may in such cases be better placed than machines to make the relevant distinctions.

There is, however, another side to this. According to some commentators, robots in certain cases conceivably are better than humans at making distinctions and can thus save the lives of civilians because of better targeting.<sup>33</sup> This could be because their sensors may be more suited to obtain or process certain kinds of information than humans, and thus make better battleground assessments. As has been alluded to, robots are also not susceptible to those human errors typically associated with factors such as fear, anger, and revenge.

To use an example, let us explore the impact of fear in a specific armed conflict scenario. A soldier who is confronted with a situation where it is not clear whether an unknown person is a combatant or a civilian, because of the instinct of survival, may shoot immediately. A ground robot, on the other hand, may utilise different tactics to go closer and, only when fired upon, return fire. In such a case, it may be easier for the robot than a human to give effect to the presumption – required by IHL – that where there is doubt as to someone’s status, they must be presumed a civilian who may not be targeted. A robot can act ‘conservatively’<sup>34</sup> and in a ‘self-sacrificing manner’ – shooting second or only firing when fired upon.<sup>35</sup> This kind of scenario may repeat itself in different circumstances, with a cumulative effect of saving great numbers of lives over time.

Autonomous weapons may find it easier, specifically, to identify status-based targets, such as members of a declared hostile force, as is often the case in international armed conflicts. In cases of narrowly-confined, low-clutter environments, the danger of autonomous weapons hitting the wrong target is also reduced. At the same time, this is not the situation that occurs most frequently. The current proliferation of asymmetric warfare and non-international armed conflicts, especially in urban environments, means that in most instances complicated assessments have to be made.<sup>36</sup> During counter-insurgency and unconventional warfare, in which combatants are often only identifiable through the interpretation of their conduct, the limited ability of autonomous weapons to interpret such open-ended data will be a significant

---

<sup>30</sup> P Asaro ‘On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanisation of Lethal Decision-making’ (2012) 94 *International Review of the Red Cross* 687, 697–99.

<sup>31</sup> See B Boothby ‘How Far Will the Law Allow Unmanned Targeting to Go?’ in D Saxon (ed) *International Humanitarian Law and the Changing Technology of War* (2013) 59.

<sup>32</sup> See Sassòli (note 27 above) p. 332.

<sup>33</sup> See note 11 above.

<sup>34</sup> See G Marchant and others ‘International Governance of Autonomous Military Robots’ (2011) XII *Columbia Science and Technology Law Review* 280.

<sup>35</sup> Singer (note 1 above) 398.

<sup>36</sup> J S Thurnher ‘Legal Implications of Fully Autonomous Targeting’ (2012) *Joint Force Quarterly* Issue 67 4th Quarter 80.



obstacle to compliance with the rule of distinction.<sup>37</sup> In some but not all cases, human judgement may thus be required to meet the requirement of distinction. However, this may change over time as technology advances.

The IHL rule of *proportionality*, for its part, requires that the expected harm to civilians – for our purposes, in the form of loss of life or the infliction of injury – be measured prior to the attack against the anticipated military advantage to be gained from the operation.<sup>38</sup> If the expected harm to civilians (the so-called incidental or collateral damage) is excessive compared to the military advantage, the attack should not proceed. Compliance with this complex rule of international humanitarian law is largely dependent on value judgements and subjective estimates of the risks, and it can be questioned to what extent robots may reliably make such assessments, even more than is the case with distinction.<sup>39</sup>

Whether an attack complies with the rule of proportionality needs to be assessed on the merits of each individual attack (seen as a whole, not each individual engagement that forms part of the attack), depending on the specific context and considering the totality of the circumstances of that attack.<sup>40</sup> The requirement of a case-by-case analysis seems to militate against the idea that a single algorithm can be made available in advance to provide all the answers, and rather call for human case by case discretion.

Proportionality generally involves and requires distinctively human and indeed subjective judgement. The prevailing legal interpretations of the rule of proportionality explicitly rely on notions such as ‘common sense’, ‘good faith’ and the ‘reasonable military commander standard’.<sup>41</sup> It is difficult to conceive that these concepts can readily be translated into computer programmes, now or in the future.<sup>42</sup>

It will be especially difficult for a computer to assess military advantage.<sup>43</sup> The value of a target in the pursuit of a military objective, which determines the level of permissible collateral damage, is constantly changing and depends, among other things, on the moment in the conflict. The open-endedness of the rule of proportionality combined with the complexity of circumstances may result in undesired and

---

<sup>37</sup> HRW (note 19 above) 30.

<sup>38</sup> Article 51 (5) (b) of Additional Protocol I to the Geneva Conventions of 1977, ‘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’.

<sup>39</sup> M Wagner ‘The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems’ (2012) <[http://robots.law.miami.edu/wp-content/uploads/2012/01/Wagner\\_Dehumanization\\_of\\_international\\_humanitarian\\_law.pdf](http://robots.law.miami.edu/wp-content/uploads/2012/01/Wagner_Dehumanization_of_international_humanitarian_law.pdf)> accessed 21 March 2017.

<sup>40</sup> P Lin, G Bekey & K Abney ‘Robots in War: Issues of Risk and Ethics’ in R Capurro & M Nagenborg (eds) *Ethics and Robotics* (2009) 57. See also Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, para 50 <http://www.icty.org/x/file/Press/nato061300.pdf> accessed 21 March 2017.

<sup>41</sup> T Hagmaier and others ‘Air Force Operations and the Law: A Guide for Air, Space and Cyber Forces’ 21, <<http://acikistihbarat.com/Dosyalar/us-air-force-operations-and-international-law-guide-acikistihbarat.pdf>> accessed 21 March 2017.

<sup>42</sup> Some proponents of the use of autonomous weapons argue that certain analyses such as proportionality would at least initially have to be made by commanders, while other aspects could be left to autonomous weapons. B Kastan ‘Autonomous Weapons Systems: A Coming Legal ‘Singularity?’ (2013) 45 *Journal of Law, Technology and Policy* 22.

<sup>43</sup> See Sassöli, (note 27 above) 331.

unexpected behaviour by autonomous weapons, and it is an open question whether technology will ever reliably cross this bridge.<sup>44</sup>

IHL, additionally, requires parties to an armed conflict to take all feasible *precautions* to limit civilian casualties.<sup>45</sup> Given their ability to loiter above a target and to use computers to work out the angle of attack that is least likely to lead to civilian casualties, it could be argued that autonomous weapons (and other unmanned systems) may meet this requirement more easily than those of distinction and proportionality.

It should also be asked whether using autonomous weapons – where the release of force occurs by definition after activation by the human into a certain environment – will allow enough opportunity to deviate from pre-set or pre-programmed plans, should an unexpected occurrence take place. An assumption of IHL seems to be that precaution will be supplemented by ‘on the spot’ human decision-making – in fact, precaution may require allowing for a last-minute change of plan. Time thus plays a crucial role.

It appears from the above data that autonomous weapons may be able to meet the requirement of distinction in only limited cases and even less so as far as proportionality is concerned. The most likely cases where they will meet the requirements of IHL arise if autonomous weapons are deployed in an area where there are no civilians present. However, this is hardly an example of machines making adequate distinction and proportionality assessments. Rather, in such cases, the distinction and proportionality decisions have been taken by the human operator who chose the terrain where the robots are deployed, serving as an exercise of human control, albeit from the ‘wider loop’. These are thus not necessarily cases of fully machine autonomy.

Using machines that employ facial recognition to identify a target also does not amount to the machine engaging in distinction – that was done by the programmer who identified the particular person as a target, and subject to other considerations such as proximity in time between the deployment of the weapon and the attack, may not qualify as a case of full machine autonomy.

The problems outlined above – that the lack of human judgement may prevent autonomous weapons from taking appropriate targeting decisions in some cases – obviously increase as human control over such decisions diminishes, suggesting that developing a notion such as ‘meaningful human control’ can help to set the cut-off line.<sup>46</sup> All of the above is subject to the consideration that targeting technology is certain to evolve over time, and may in the process diminish the need for human involvement in order to be accurate. Moreover, to the extent that using technology can assist better targeting outcomes, there may be a duty on commanders to use it.<sup>47</sup>

---

<sup>44</sup> N Sharkey ‘Automated Killers and the Computing Profession’ (2007) 40 *Computer* 122; A Krishnan *Killer Robots: Legality and Ethicality of Autonomous weapons* (2009) 98–9.

<sup>45</sup> Article 57 of Additional Protocol I to the Geneva Conventions: ‘(1) In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects’. See also J D Herbach ‘Into the Caves of Steel: Precaution, Cognition and Robotic Weapon Systems under the International Law of Armed Conflict’ (2012) 4 *Amsterdam Law Forum* 30.

<sup>46</sup> On this concept, see Section 4 below.

<sup>47</sup> For an analogous argument concerning drones, see B Strawser (ed) *Killing by Remote Control: The Ethics of an Unmanned Military* (2013) 17. See also M C Horowitz & P Scharre ‘Do Killer Robots Save Lives?’ (2014) <<http://www.politico.com/magazine/story/2014/11/killer-robots-save-lives-113010.html>> accessed 21 March 2017.

It thus seems fair to say that the number of instances where the outcome of machine targeting will comply with the IHL rules on targeting is bound to increase over time, as technology advances.

The question whether autonomous weapons are able to do proper targeting brings to the fore the further issue of the yardstick against which their performance is to be measured. Some authors have argued that to be acceptable, autonomous weapons will be required merely to match the performance of the average human soldier.<sup>48</sup> This may require further consideration. Higher levels of control imply higher levels of responsibility. To be considered potentially acceptable, autonomous weapons will in my view at least have to withstand closer scrutiny than the average human soldier, even if the objective standards remain the same.<sup>49</sup>

This brings us to accountability. From a human rights perspective, the right to life is violated not only when a particular deprivation of life is arbitrary, but also when there is an absence of proper accountability.<sup>50</sup> Accountability in respect of war crimes is fully established in IHL, and there is increasing focus on accountability for other violations of international law.<sup>51</sup> Even if accountability is often not achieved in practice, it is important as far as ensuring the recognition of the norms is concerned that it is possible in principle.

If autonomous weapons outperform humans as far as targeting is concerned, there will be fewer mistakes, but there will still be mistakes.<sup>52</sup> The question is who will assume legal and moral responsibility for those mistakes. It has been argued that autonomous weapons leave an accountability gap, and also from that perspective can

---

<sup>48</sup> See note 41 above 50; Sassòli (note 28 above) 320.

<sup>49</sup> According to the African Commission on Human and Peoples' Rights *General Comment 3 on the Right to Life*: 'The use during hostilities of new weapons technologies ... should only be envisaged if they strengthen the protection of the right to life of those affected'. (2015) para 35 <<http://www.achpr.org/instruments/general-comments-right-to-life>> accessed 21 March 2017. In many areas the law expects higher levels of expertise to be accompanied by higher levels or standards of conduct. Medical specialists are expected to have a greater level of skill than general practitioners. See Legal Match 'Medical Specialist Malpractice' <<http://www.legalmatch.com/law-library/article/medical-specialist-malpractice.html>> accessed 21 March 2017. We hold those who are responsible for air travel to higher safety standards than is the case with road travel. See, eg 'Difference Engine: Up, Up and Away', *The Economist* (7 January 2013) <<http://www.economist.com/blogs/babbage/2013/01/air-safety>> accessed 21 March 2017. It seems that we are, in short, more tolerant of human than technological failings; more accepting of human rather than machine unpredictability. Experiments in human psychology indicate – and the public reaction against the possible introduction of autonomous weapons seems to confirm – that there will be a stronger outcry if the use of algorithms result in atrocities than if those atrocities were caused by humans. D Kahneman *Thinking, Fast and Slow* (2011) 229. Armed drones have attracted levels of international scrutiny which arguably would not have been the case if manned systems were used. This may suggest emerging state practice to hold such technologies to a higher standard, or at least higher levels of scrutiny. Requiring human-replacing robots merely to match humans as far as accuracy in targeting is concerned is also problematic because humans who err or transgress in battle often are treated with high levels of tolerance. Those who judge soldiers accused of wrongdoing or decide whether they should be prosecuted are in most cases their peers – their 'brothers in arms' – and are bound to empathise and identify with them. Holding autonomous weapons to the standards applicable to humans would entail expanding the application of those low standards.

<sup>50</sup> See footnote 22 above.

<sup>51</sup> See the United Nations Office of the High Commissioner for Human Rights, Revision of the UN Manual on the Effective Prevention and Investigation of Extra-Legal, Arbitrary and Summary Executions (the Minnesota Protocol) <<http://www.ohchr.org/EN/Issues/Executions/Pages/RevisionoftheUNManualPreventionExtraLegalArbitrary.aspx>> accessed 21 March 2017.

<sup>52</sup> See G Veruggio & K Abney 'Roboethics: The Applied Ethics for a New Science' in P Lin, K Abney & G A Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics* (2012) 114; R Sparrow 'Killer Robots' (2007) 24 *Journal of Applied Philosophy* 62.

violate the right to life. Much has been written about the topic, and a few general comments will have to suffice here.<sup>53</sup>

Who is legally responsible and accountable if autonomous machines use force in a way which would normally be considered a war crime or another breach of international law when, due to a malfunction or unforeseen circumstances, the wrong target is struck or there are excessive civilian casualties? If autonomous weapons indeed take over the role of the warrior, the problem is that beyond a certain point of machine autonomy this warrior cannot, in a meaningful way, be held responsible for its actions. With human decision-making, it makes sense to talk about right and wrong; to the extent that decisions are made exclusively by machines the question is confined to whether they are effective or not.

An important basis of accountability is the exercise of control over a situation by the person being held responsible. However, even though autonomous weapons are by definition autonomous, they are not moral agents, and putting a robot in jail or requiring it to pay a fine is an obvious exercise in futility. Traditionally, accountability is strictly a human affair.<sup>54</sup>

If it is correct that accountability is premised on some level of control, it is hard to see how there can be responsibility without meaningful human control. Conversely, if there is meaningful human control, even if that occurs at the outer edges of the decision-making loop, those who exercised it can be identified and if needed, blame can be apportioned according to their level of control. At the same time, the practical difficulties of assigning responsibility on multiple levels must be kept in mind – if everyone is responsible, in effect no one is.

In a world dominated by computers and where power is legitimised by algorithms, the logical response to errors on the battlefield may be to call for technical improvements rather than traditional concepts of human accountability. Given the fact that accountability is an element of the protection of the right to life, this would entail a fundamental change in our understanding of the right to life.

### **3. Should they do it? Robots and the power to determine who will live or die**

Most discussions of autonomous weapons address some articulation of the first question – ‘Can they do it?’ and leave it at that. The starting point seems to be that if robots can at least match human targeting on a consistent basis (and provided a

---

<sup>53</sup> See, eg R Crootof ‘War Torts: Accountability for Autonomous Weapons’ (2016) *University of Pennsylvania Law Review* 1347; K Cass ‘Autonomous Weapons and Accountability: Seeking Solutions in the Law of War’ (2015) 48 *Loyola of Los Angeles Law Review* 1016; D N Hammond ‘Autonomous Weapons and the Problem of State Accountability’ (2015) *Chicago Journal of International Law* 651; J I Walsh, ‘Political Accountability and Autonomous Weapons’ (2015) *SAGE* 1.

<sup>54</sup> The Nuremberg Tribunal has famously held that ‘Crimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced’. See Office of United States, Chief of Counsel for Prosecution of Axis Criminality ‘NAZI Conspiracy and Aggression Opinion and Judgment’ (United States Government Printing Office, 1947) 53.

solution to the legal accountability issue can be found), they should be allowed; but if that this is not the case their use should not be permitted. People who follow this approach may thus have what we can call a pragmatic objection to (at least some kinds of) autonomous weapons. They may, for example, believe that technology will never be able to match human targeting on a consistent basis, or law cannot be reformed to ensure appropriate accountability mechanisms are in place (assuming accountability remains a requirement for the protection of the right to life). Seen from this perspective, the requirement of meaningful human control is needed not as a matter of principle, but in order to secure accuracy in targeting and as a basis for legal reform. Clearly, this approach is conditional on technological developments, and may or not be trumped depending on such progress.

From a different perspective, however, the ‘can they do it?’ question is not decisive. Even if the answer to that question were to be yes, and in some cases autonomous weapons can match or surpass human targeting, or they will be able to do so in the future when technology is further developed, the principled question still has to be answered: Is it not inherently wrong – as a matter of morality and of law – to let machines decide who should live and who should die? If the answer to the second question is negative, autonomous weapons should be prohibited, even if it can conceivably lead to better targeting.

While the discussion of the ‘can they do it?’ question dealt with the plight of civilians and others *who are protected* from direct or indirect targeting, the ‘should they do it?’ question also brings into play the deployment of autonomous weapons against *those who are not protected*, and who can thus be legitimate direct targets or incidental casualties in terms of the standard rules of IHL. The issue in these latter cases is not the *accuracy* of the targeting – it is assumed that a legitimate direct or indirect target is hit. The enquiry is thus shifted to the *manner* of targeting, namely through a fully autonomous robot. All else being equal, is it of ethical or legal importance whether a human or an algorithm makes the determination that a missile will be launched against a human being?

One often hears the argument that it makes no difference whether a human or a machine released a missile that hit someone – the result is the same, namely death, and there should be no special legal consequences as long as the requirements of IHL are otherwise complied with. There is often little sympathy in this context for those who are at the receiving end of force if the law provides that they may as a result of their status or conduct be directly targeted (e.g. enemy combatants or civilians directly participating in hostilities).

I want to argue that there are right to life as well as right to dignity reasons why full machine autonomy over critical functions in using force against those who may otherwise be targeted should not be permitted. Even if someone may be killed, that does not mean that person may be killed by fully autonomous machines.

As noted earlier, the test whether the right to life has been violated is whether the deprivation of life was ‘arbitrary’. Is fully autonomous targeting inherently ‘arbitrary’? Philosopher Peter Asaro has argued that an implicit requirement is to be found in IHL that lethal force may only be used based on a deliberative human decision, and that making the determination to use such force cannot be delegated to an

autonomous machine.<sup>55</sup> One is not allowed to delegate a power that one does not have – and humans do not have the authority to use force without applying their minds. Non-human decision-making, or determinations where there is no meaningful human control regarding the use of lethal force is, according to this argument, inherently ‘arbitrary’, and deaths that result are unlawful deprivations of life.<sup>56</sup>

Much of our practice seems indeed to be premised on the idea that, where people’s lives or other grave consequences are at stake, a human – someone identifiable – should apply their mind and assume ultimate responsibility. Put differently, there should be ‘deliberative human decision-making’ or – what may amount to the same thing – ‘meaningful human control’. Moreover, there must potentially be accountability.

As noted earlier, robots already affect our lives profoundly. This includes their impact on life-and-death issues.<sup>57</sup> Self-driving cars will have to make determinations that affect life – for example, should they swerve out to avoid hitting a number of people if that means running over a single pedestrian on a sidewalk? Yet, the case of self-driving cars inflicting such incidental harm, which it is designed to avoid, is very different from the case of weapons that are specifically designed and used to inflict deadly force.

Our moral practices in other areas also seem to point towards great caution. For example, the idea that computers should serve as judges in courts of law, in spite of initial enthusiasm when computers were first developed, does not seem to attract many serious supporters today, in spite of all the advances in technology. This is the case especially in as far as criminal law is concerned, given its far-reaching impact on the human person.<sup>58</sup> The notion that computers can assume final ‘responsibility’ to impose the death penalty or even sentences involving deprivation of freedom is unthinkable, even in the most authoritarian regime.

One reason why we do not accept computers as judges may be that computers ‘can’t do it’ – that they are technically incapable of making such judgements. However, I would argue that even if they could pass something similar to the Turing test – if lawyers cannot distinguish the rulings made by computers who are programmed to exercise judicial functions from decisions made by human judges – I would venture to say that we still see justice as something that cannot be dispensed by computers. Clearly, judges can and should use computers to inform their decisions, but in the end, they have to make the value calls. Justice is ultimately a strictly human affair. Decision-making in the context of courts is different from decision-making in war, in that the latter entails a response to an emergency situation. However, it seems that there is something disconcerting about the notion of computerised and thus

---

<sup>55</sup> Asaro (note 16 above) 689. The Martens Clause, a longstanding and binding rule of IHL, specifically demands the application of ‘the principle of humanity’ in armed conflict. Article 1(2) of Additional Protocol I to the Geneva Conventions. See also the preambles to the 1899 and 1907 Hague Conventions as well as of the Hague Convention with Respect to the Laws and Customs of War on Land and its Annex: Regulation Concerning the Laws and Customs of War on Land (Hague Convention II). See T D Evans, ‘At War with the Robots: Autonomous Weapon Systems and the Martens Clause’ (2014) 41 *Hofstra Law Review* 712–14.

<sup>56</sup> Asaro (note 30 above) 699.

<sup>57</sup> See ‘Surgical Robots’ <<http://www.allaboutroboticsurgery.com/surgicalrobots.html>> accessed 21 March 2017.

<sup>58</sup> An early – and controversial – proposal in this regard was made in 1991 by HJ van den Herik. See A Schmidt ‘Ought Computers Adjudicate?’ (2007) *MICC Universiteit Maastricht* 133–47. See, however, Stanford University’s ‘Smart Prosecution’ Project <<https://law.stanford.edu/projects/smart-prosecution/>> accessed 21 March 2017.

anonymous determination over vital aspects of peoples' lives in both cases, especially where life and death is at stake.

One way of interpreting the prohibition on the arbitrary deprivation of life is to argue that the involvement of computers in targeting decisions, given their quantitative approach, is the surest way to avoid arbitrariness. Computers, it could be argued, offer an objective as opposed to an invariably subjective way of determining how to direct force.

However, as seen earlier, autonomous weapons are to some extent unpredictable. While computers may be able to respond in a consistent way when presented with the same stimuli, the complex environment of war seldom presents itself twice in exactly the same detail. As a consequence, reality simply may be too complex for computers to do justice to the full range of nuances of every case. Computer consistency may just be too blunt for the complexity of life.<sup>59</sup> Moreover, machine learning may result in different results over time, by different robots. Artificial intelligence, irrespective of highly developed it is, provides no guarantee of achieving the Truth, in the sense of bringing us to the one and only right answer.

The decision to use force against a human being, even if done in the heat of battle, requires at some level – somewhere in the decision-making loop that results in an attack, not necessarily as far as every individual release of force is concerned – an element of human deliberation.<sup>60</sup> Someone must take the time and apply his or her mind and 'sign off' on the conclusion that using force is really necessary as far as possible *in this particular case*.<sup>61</sup> Philosopher Thomas Nagel has emphasised in a different context that the use of force on the battlefield, to be potentially justifiable, must remain a manifestation of inter-personal relationships.<sup>62</sup> In the words of Hannah Arendt in a different context: 'Particular questions must receive particular answers'.<sup>63</sup> The decision to use deadly force, every time it is made, entails crossing a new threshold because of the unique value of each individual life. I have argued elsewhere that war without on-going human reflection is mechanical slaughter.<sup>64</sup>

The greater use of unmanned systems and in particular autonomy in the critical functions implies the (further) dehumanisation of lethal decision making.<sup>65</sup> Machines lack morality and mortality, and they should as a result not be given life-and-death powers over humans. Machines have no understanding of the importance of life, and of the implications of taking it. While some humans also lack these qualities, machines by their very nature can never have it. Cold-blooded – or rather bloodless – entities should not have the ability to take decisions on whether to spill human blood.

---

<sup>59</sup> Consistency, in itself, is not necessarily a virtue. Many forms of killing, such as genocide, are consistent, but this does not diminish the blameworthiness the perpetrators – in fact, it increases it.

<sup>60</sup> See Asaro (note 30 above) 695.

<sup>61</sup> On the need to take time with decisions involving trade-offs between 'sacred values' such as life, see P E Tetlock 'Thinking the Unthinkable: Sacred Values and Taboo Cognitions' (2003) 7 *Trends in Cognitive Sciences* 320, 322.

<sup>62</sup> T Nagel 'War and Massacre' (1972) 1 *Philosophy and Public Affairs* 123.

<sup>63</sup> See H Arendt *Responsibility and Judgement* (2009) vii.

<sup>64</sup> See 'UN Human Rights Expert Urges Global Pause in Creation of Robots with 'Power to Kill'' UN News Centre (30 May 2013) <<http://www.un.org/apps/news/story.asp?NewsID=45042#.WJruGxJ97Jw>> accessed 21 March 2017.

<sup>65</sup> See Asaro (note 30 above) 687. Also see M Wagner 'Autonomy in the Battle Space: Independently Operating Weapon Systems and the Law of Armed Conflict' in D Saxon (ed.), *International Humanitarian Law and the Changing Technology of War* (2013). See also HRW, *Losing Humanity* (note 19 above) 2.

The concept of time seems to play a central role in what we regard as an arbitrary taking of life. Respect for the value of life is most clearly manifested by the requirement of human rights law to leave the decision whether someone has to be killed to the last moment possible when there is no other alternative to save another life. IHL does not demand that the use of force is justified only as an *ultimo ratio* against an 'imminent' threat, but that does not mean that time or some level of urgency does not play any role in armed conflict.

War is and must remain a temporary exception, based on an ongoing emergency. Traditionally, the presence of a soldier on the battlefield to personally take the decision whether it is, in the final analysis, necessary to end someone's life (often with the soldier as a result also accepting personal risk) is to some extent a manifestation that such an emergency exists, and a manifestation of respect for the longer term value of life. It could be argued that only for as long as you place yourself at risk can you do it to others. This element is compromised when humans are replaced by any unmanned system, and it is absent when they are replaced by robots.<sup>66</sup>

Developing autonomous weapons involves taking the decision to kill people in advance, in the abstract, and based on hypothetical scenarios. When a deadly algorithm is developed, or a fully autonomous weapon activated, the person who does this can by definition not have the full picture in mind which will apply when force is eventually released. He or she clearly also do not place themselves at risk.

An analogous case is that of the mandatory death penalty. It is widely seen as a violation of the right to life for a legislature to rule in advance, in a generic law, that judges have to impose the death penalty on all accused who meet certain criteria. If the death penalty should be used at all, it must be left to a judge in a concrete case to decide that there is no other alternative punishment that could be imposed.<sup>67</sup> Autonomous weapons are set up to respond in a generic way to its environment – and does not entail a consideration of the unique circumstances of a specific, concrete case.

The same problem underlies the famous ticking bomb scenario which purportedly shows that torture will become acceptable if the lives of enough people depends on torturing someone who may have the information necessary to diffuse a time bomb.<sup>68</sup> Playing along with this scenario puts one on the slippery slope towards accepting that the prohibition against torture is not a general principle. The ticking bomb scenario does not prove torture is acceptable because it asks us to take a position based on a hypothetical case in the abstract, where a real emergency situation has not presented itself, much as is the case with those who programme autonomous weapons.

The argument that autonomous weapons may not be used because the absence of deliberative human decision-making in the case at hand will violate the right to life of those targeted is a principled objection, which does not depend on the accuracy of the targeting. However, what is the foundation of this approach – what is the deeper value that is compromised when autonomous targeting takes place?

---

<sup>66</sup> See P W Kahn, 'The Paradox of Riskless Warfare' (2002) 22 *Philosophy and Public Policy* 1–3; P W Kahn 'War and Sacrifice in Kosovo' (1999) 19 *Philosophy and Public Policy* 1.

<sup>67</sup> See, eg UN Human Rights Committee CCPR/C/70/D/806/1998 para. 8.2

<sup>68</sup> M Sandel *Justice: What is the Right Thing to Do?* (2009) 38–40.



In my view, what is at stake is not the protection of mere, or ‘bare life’,<sup>69</sup> or the continuation of biological existence. Something much more meaningful is potentially undermined by autonomous targeting, namely the protection of dignified life.<sup>70</sup> It is dignity that gives life its meaning. The right to life and the right to dignity cannot be divorced – they are, in the standard language, interrelated and indivisible. Even if there is a legitimate limitation on someone’s right to life, as is the case where the person is not protected under IHL, that person retains his or her other rights, such as the right to dignity. It seems worth exploring more specifically what the dignity component of the right to life brings to the discussion in the context of autonomous weapons, again focusing in particular on the use of autonomous weapons against those who may otherwise be targeted.

Human rights in general, and dignity in particular, emphasises the notion that each person is entitled to be treated according to his or her own full merits. In the words of Ronald Dworkin, the concept of human rights requires that each person is entitled to ‘equal concern and respect’.<sup>71</sup> This means we should not simply be treated equally but also need to be taken seriously as separate and irreplaceable individuals.

As alluded to above, the concept of dignity underlies human rights law as well as IHL. However, many international human rights treaties do not recognise dignity as an explicit right. While the Universal Declaration of Human Rights does recognise this right, the International Covenant of Civil and Political Rights makes passing reference only to dignity,<sup>72</sup> and the European Convention on Human Rights does not recognise it as a right.

By contrast, this right is clearly articulated in the African Charter on Human and Peoples’ Rights, which provides that ‘Every individual shall have the right to the respect of the dignity inherent in a human being and to the recognition of his legal status. All forms of ... inhuman or degrading punishment and treatment shall be prohibited’.<sup>73</sup> The African Commission of Human and Peoples’ Rights, in their General Comment 3 on the right to life, have emphasised the interrelated nature of the right to life and the right to dignity, and have thus stated that what is being protected is the right to a ‘dignified life’.<sup>74</sup>

The notion of dignity is directly challenged by the idea that people, including opponents in war, can become a casualty of the algorithmic calculations (literally reducing its targets to the 0’s and 1’s of the digital code) of an unthinking entity if we happen to be in its way. Where that occurs, one person’s death is indistinguishable from that of so many others who happen to find themselves in the striking range of

---

<sup>69</sup> G Agamben *Homo Sacer Sovereign Power and Bare Life* (1998).

<sup>70</sup> See *Gyan Kour v State of Punjab*, AIR 1996 SCC 946; A Jakubowski, *Cultural Rights as Collective Rights: An International Law Perspective* (2016) 238; J K Das, *Human Rights Law and Practice* (2016) 231.

<sup>71</sup> R Dworkin, *Taking Rights Seriously* (1978) xii.

<sup>72</sup> Article 10, dealing with the rights of those deprived of their liberty.

<sup>73</sup> Article 7.

<sup>74</sup> See paras 3 and 6 of the African Commission on Human and Peoples’ Rights *General Comment 3 on the Right to Life* (note 49 above). The broad understanding of the right to life that includes among other things dignity is based according to the Commission on ‘widely shared communal values of the continent, according to which the value of one person’s life is tied to the value of the life of others’ (para 3). T Chengeta ‘Dignity, Ubuntu, Humanity and Autonomous Weapon Systems (AWS) Debate: An African Perspective’ (2016) 13 *Brazilian Journal of International Law* 460–502 also links dignity to *Ubuntu*, the Nguni concept that ‘I am because we are’. The right to dignity plays an integral part in the constitutional jurisprudence of South Africa (section 10). See, eg L Ackerman *Human Dignity: Loadstar for Equality in South Africa* (2012).

generic killing machines. Restrictions on the use of force are not simply a numbers game – it is also about protecting the value of each individual life.

A central thrust of the notion of human dignity is the idea that humans should not be treated as something similar to an object that simply has an instrumental value (as is the case e.g. with slavery or rape) or no value at all (as with many massacres).<sup>75</sup> The person against whom the force is directed by autonomous weapons is reduced to being an object that has to be destroyed, and that is even more clearly the case where incidental casualties are at stake. They have no avenue, futile or not, of appealing to the humanity of the enemy, or hoping their humanity will play a role because it is a machine on the other side.

The use of terms such as ‘killer robots’ to denote fully autonomous weapons may well be criticised. However, the strength of the intuitive reactions that the use of autonomous weapons is likely to elicit cannot be ignored, especially if it is recognised that law is not simply a matter of logic and rationality, but of the protection of human values and deeply engrained psychological concerns.

The realities of modern warfare are, of course, such that someone who is about to be targeted in many cases does not have a real chance of appealing to the humanity of the person on the other side. However, the hope that this may be possible has so far not been completely excluded. With the introduction of autonomous weapons, there is no such chance. Having autonomous weapons as a legal and legitimate part of the world in which we live can undermine an important part of our hard-wiring: namely, hope. Psychologists have long emphasised the importance of a measure of optimism in human beings.<sup>76</sup> An emphasis on hope indeed underlies many religious and other world views, as well as philosophical traditions.<sup>77</sup> For machines to have the power of life and death over human beings may change some of our deep-seated assumptions about the world in which we live, and the extent to which we experience it as a place in which empathy, redemption, and mercy may be found. Where it is legally or even ethically permissible to use force, humans may decide not to do so because something holds them back. This possibility is excluded when autonomous weapons are used.<sup>78</sup> This is the classical situation where a decision-maker says ‘It’s lawful, but it’s awful’ – and as a result not to do it.

If autonomous weapons can be calibrated to use force to the full extent that the law allows humans (as the forces at work at weapons markets may dictate), lives that could have been spared will be lost because, to the extent that the human is taken out of the decision-making process, so is the potentially restraining elements of humanity. In a pre-autonomous weapons world, there are two thresholds that force have to meet during armed conflict. The first is that it has to be established that the use of

---

<sup>75</sup> C McCrudden (ed) *Understanding Human Dignity* (2014).

<sup>76</sup> See T Sharot *The Optimism Bias: A Tour of the Irrationally Positive Brain* (2011).

<sup>77</sup> For example, M Heidegger as a central figure in existentialism is famous for insisting on the priority of the possible over the actual. See also E Bloch *The Principle of Hope* (1981).

<sup>78</sup> An extraordinary example is provided by the case of Franz Stigler. In December 1943, Lt Franz Stigler, a German pilot, enabled a damaged Allied B-17 plane to land. To prevent further fire from German planes, Stigler escorted the B-17 to safety, despite having an opportunity to shoot it down which would have qualified him for a Knight’s Cross. See T Rennel ‘A Stricken Allied Bomber, the German Ace Sent to Shoot it Down and a Truly Awe Inspiring Story of Wartime Chivalry’ (2013) <<http://www.dailymail.co.uk/news/article-2370933/A-Higher-Call-A-stricken-Allied-bomber-German-ace-sent-shoot-truly-awe-inspiring-story-wartime-chivalry.html>> accessed 21 March 2017.

force in a particular attack is legal. The second is that a human being moreover must deem it necessary and warranted in that particular case to use such force. Autonomous weapons mean the second barrier is removed.

Some would argue that it no longer matters to the person concerned whether he or she was killed by a missile launched by a human or a machine. However, knowing that it may happen in future that one is directly targeted or becomes collateral damage, I would suggest, affects the dignity of each person in that situation.

There is a difference between dying a 'good death' and dying a 'bad death', and this difference affects the meaning we attach to human life.<sup>79</sup> As Harvard medical doctor Atul Gawande has emphasised, human life is best understood as a narrative, and in stories, the end matters.<sup>80</sup> An end at the hand of machine differs from an end at the hand of humans. This concern affects the value of human life *per se*. Fully autonomous weapons will make the world a bleaker place. This is supported by the fact that IHL recognises the notion of the dignity of the dead.<sup>81</sup>

Dignity can be approached from different angles. There is the dignity of the person at the receiving end, under discussion above, who is reduced to being a target or incidental damage. But the autonomous use of force can also implicate the dignity of the people on the side that deploys autonomous weapons because it is being done in their name. Dignity entails the ability to be a moral agent: to be autonomous in the sense that they exercise moral choices and to be responsible for their outcomes. Fully autonomous weapons as a human-replacing technology may remove this possibility to be a moral person. Conversely, where there is meaningful human control, autonomous weapons can potentially increase human autonomy and control over the outcomes of one's actions, and thus arguably enhance human dignity. There is little sense in arguing that the right to life is 'the supreme right', and as a result, the prospect that autonomous weapons may save lives through better targeting outweighs considerations of dignity, if what is protected is not 'bare life' but dignified life.

The asymmetric nature of a battle where one side is represented by deadly robots and the other side is present as mortal human beings, also has a special resonance for concerns of dignity, which will be exacerbated where these systems are used by the developed world against the developing world. Also from this point of view it will be important for regions such as Africa to consider its interests in the development of this new technology and to engage fully with the global debate and in helping to shape what a dignified life means.

It is instructive to consider the ruling of the German Constitutional Court that legislation allowing the Minister of Defence to authorise the shooting down of a

---

<sup>79</sup> Even if it is assumed that the death penalty in a particular case is not in violation of international law, certain methods of execution, such as beheading, stoning and execution by gas asphyxiation are explicitly prohibited. See A/67/279, Interim Report of the Special Rapporteur on Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, J Mendez (9 August 2012) paras 6–8 <[http://www.un.org/ga/search/view\\_doc.asp?symbol=A/67/279](http://www.un.org/ga/search/view_doc.asp?symbol=A/67/279)> accessed 21 March 2017. However, it should be recognised that in the cases of armed conflict the method of death remains the same whether a human or a computer releases the force (eg a rocket is launched); what differs is whether a machine or a human being determines that such force will be deployed. See on the concept of a dignified death, D Kahneman (note 50 above) 140.

<sup>80</sup> See A Gawande *Being Mortal: Illness, Medicine and What Matters in the End* (2014).

<sup>81</sup> See rule 113 of the ICRC Customary Law Database <[https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_chapter35\\_rule113](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_chapter35_rule113)> accessed 21 March 2017.

civilian aircraft involved in a 9/11 style terror attack was unconstitutional.<sup>82</sup> This is in spite of the lives of those that potentially would be saved (in the 9/11 scenario, those in the building), *inter alia* based on the argument that shooting down the plane would constitute a violation of the right to dignity of those on board.<sup>83</sup> Part of what is wrong with such legislation is that it involves the legislature in deciding in a hypothetical case, in advance and without being compelled by the circumstances to accept that there is no other alternative, that a certain category of people may be killed. It is the same problem that we encountered with the mandatory death penalty and the ticking bomb scenario and torture, and with programming fully autonomous weapons.

This brings us once more to the issue of accountability, but in this case not necessarily legal accountability. Because the concern of the ‘should they?’ question is with the fate of those who are not protected by IHL from attack, directly or indirectly, the traditional issues of legal accountability for the possible violations of the right to life – seen as biological life – do not arise in this part of the discussion.<sup>84</sup> However, the introduction of autonomous weapons raises novel and important issues of moral and political accountability, and indeed if it is accepted that the right to life protects dignified life, legal accountability may also arise.

The fact that each life has value, and the right to life is seen as the right to dignified life, presupposes that someone, somewhere will internalise the cost of crossing the threshold of ending it; a human being or beings can and in many cases will assume or be assigned moral or political responsibility for its loss, even if it is otherwise legally permitted. In common parlance, taking a life under any circumstances rest on conscience of the person responsible.

International law is not pacifist. War is sometimes a necessary evil. However, there is a cost, even if it is legal to kill someone. This cost has traditionally been passed on through generations by those who know war, from parents to their children and grand-children; it is integrated in thousands of ways in our various cultures. We have a moral compass, which may lose its bearings at times, but which has the potential to correct itself.<sup>85</sup>

On a very fundamental level, most people recoil from killing, and will go to great lengths to avoid it. ‘To prevent the scourge of war’ was indeed one of the main objectives of the battle weary leaders who gathered to form the UN.<sup>86</sup> In spite of the laments of our singers and philosophers that we do not learn from our mistakes in war – and probably in no little measure due to their laments – we actually sometimes do. In an era of machine learning there is a danger of overlooking the need to allow room for and to depend on human learning.

---

<sup>82</sup> German Federal Constitutional Court case on legislation authorising the shooting down of an aeroplane, 1 B v R 357/05, 15 February 2006.

<sup>83</sup> For a critical view of this case, see G Nolte ‘The Bundesverfassungsgericht on the German Aerial Security Law: A Sonderweg from the Perspective of International Law?’ in C Tomuschat, E Lagrange & S Oeter (eds) *The Right to Life* (2010) 83–94.

<sup>84</sup> See, however, *Hamiyet Kaplan and Others v Turkey* (application no. 36749/97) 2005 ECHR.

<sup>85</sup> Indeed, researchers have registered an overall long term drop in violent death globally, including death due to war, over the last four centuries and probably much longer. See S Pinker *The Better Angels of our Nature: The Decline of Violence in History and its Causes* (2011).

<sup>86</sup> C Heyns ‘The Preamble of the United Nations Charter: The Contribution of Jan Smuts’ (1995) 7 *African Journal of International and Comparative Law* 329–48.

It is difficult to see why the inner logic of machine learning over the decades would tend in the same direction, given that machines have no inherent ability to distinguish between good and evil, even if some such notions are initially programmed into some of the prototypes. The capacity to distinguish right and wrong is not part of their very nature. There is little reason to believe that machines will try to find creative ways over time to avoid violence altogether. This, to put it bluntly, is not in their culture or DNA.

Many discussions about autonomous weapons focus on the likelihood of compliance with legal or ethical standards in individual cases only, and not on their collective effect over time. If humanity now decides to allow fully autonomous weapons, it is certain to stay with us for the centuries to come. To the extent that autonomous weapons are going to do the killing in our future wars, and humans increasingly defer to machines, there will be fewer incentives for individual human beings to work to prevent war as far as possible from occurring in the first place.

What is required therefore is not only meaningful human control over every individual attack, and arguably that is not the only or even main application of the term. The main concern animating the reaction against the idea of full machine autonomy is rather the prospect of losing meaningful human control over the long-term use of algorithms doing targeting in war in general. We may get so carried away by the short term and individualised advantages of machine targeting that we do not realise that we have lost control over the enterprise as a whole before it is too late. Preventing this can only be done by adopting a new legal norm.<sup>87</sup>

#### **4. The concept of 'meaningful human control'**

UK-based NGO Article 36, drawing on the IHL principle of humanity, first proposed that the standard for the acceptability of autonomous weapons be the exercise of 'meaningful human control' over each individual attack.<sup>88</sup> Autonomous weapons that do not meet that standard – in other words, fully autonomous robots – should be banned.<sup>89</sup> This concept has been taken up widely in the international debate on AWS. For example, the General Comment on the right to life of the African Commission, provides that 'Any machine autonomy in the selection of human targets or the use of

---

<sup>87</sup> It should be noted that establishing whether there has been a breach of IHL has to be considered in respect of each separate attack – they are not to be measured in the aggregate. See Andrew Clapham and others stating that in terms of IHL, there is a need 'to distinguish between the legality of use of force at a macroscopic level and the legality of each singular military action during the conflict to which of course relevant rules of international humanitarian law shall apply'. A Clapham & Paola Gaeta *The Oxford Handbook of International Law in Armed Conflict* (2014) 555.

<sup>88</sup> See remarks by T Nash, Director, Article 36, Informal Expert Meeting on Lethal Autonomous Weapon Systems, Convention on Certain Convention of Weapons, Geneva (2014) <[http://www.unog.ch/80256EDD006B8954/%28httpAssets%29/26033D398111B4E8C1257CE000395BBB/\\$file/Article36\\_Legal+Aspects\\_IHL.pdf](http://www.unog.ch/80256EDD006B8954/%28httpAssets%29/26033D398111B4E8C1257CE000395BBB/$file/Article36_Legal+Aspects_IHL.pdf)>; See also Article 36, 'The United Kingdom and Autonomous Weapons Systems' (2016) <<http://www.article36.org/wp-content/uploads/2016/04/UK-and-LAWS.pdf>> accessed 21 March 2017.

<sup>89</sup> See note 18 above.

force should be subject to meaningful human control'.<sup>90</sup> However, there is still great uncertainty about what the concept entails.<sup>91</sup> Again, a brief discussion will have to do.

Making the litmus test for the acceptability of machine autonomy some notion of 'human control' is attractive, because the main reasons why autonomous weapons potentially could violate the rights to life and to dignity outlined above relate to the element of human control. Human judgement, for example, is called for in proportionality decisions, and to ensure that an element of deliberation is at least potentially possible where decisions about taking life are concerned. Dignity is potentially affected where a sufficient level of human control is absent or weak. As we saw earlier, control is also closely tied to and, in fact, a precondition for responsibility. However, how much human control is needed to ensure that it is 'meaningful'?

In essence, the requirement of meaningful human control demands that autonomous weapons remain *tools* in the hands of humans – do they complement or substitute human decision-making? Spatial as well as temporal proximity clearly plays a crucial role.

Meaningful human control clearly does not mean 'full' human control, or even that humans exercise stronger or more decisive control than machines. One can even foresee that the requirement of meaningful human control is met if the role of humans is limited to the original programming of the robot and humans are given an opportunity to override the machine's determination – as long as the power to override is a real power, and one that is not rendered illusory because in practice everything goes so fast that there is no room for human decision-making.

Meaningful human control could be seen as the flipside of the coin of accountability: is there enough human control over the release of force to allow for effective accountability? Requiring the presence of meaningful human control in weapons delivery has the potential to address the issues raised by the two rights under discussion when we considered the questions 'can they do it?' and 'should they do it?'. Where sufficient human control is present, the release of force will arguably be subjected to 'deliberative human decision-making', which as we saw may be an implicit requirement of the protection of the right to life. Accountability is also possible where humans exercise sufficient levels of control. Likewise, the problem of machines reducing humans to objects that need to be destroyed, and thus infringing their dignity, also does not arise where the control of humans in the decision-making process is meaningful.

The concept of meaningful human control was developed to apply to specific attacks, which is in line with the IHL approach to focus on particular attacks. However, it may also have broader significance, in particular to policy decisions

---

<sup>90</sup> See para 32 of African Commission on Human and Peoples' Rights *General Comment* 3.

<sup>91</sup> See UNOG, '2015 Meeting of Experts on LAWS' <<http://bit.ly/1jSiCro>> accessed 21 March 2017. See also Article 36 'Key Areas for Debate on Autonomous Weapons Systems' (2014) 2 Briefing Paper <<http://www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf>>. For a widely used recent exposition, see UNIDIR 'The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control Might Move the Discussion Forward' UNIDIR Resources No 2 (2014) <<http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf>> accessed 21 March 2017. Report of the ICRC expert meeting on 'Autonomous Weapons Systems': Technical, Military Legal and Humanitarian Aspects' (2014) <<https://www.icrc.org/eng/assets/files/2014/expert-meeting-autonomous-weapons-icrc-report-2014-05-09.pdf>> accessed 21 March 2017.

where decisions have to be taken about possible general rules allowing autonomous weapons or not. It may be dangerous to allow the use of certain autonomous weapons in certain circumstances, to advance military as well as humanitarian objectives, if they can be abused in other cases. That is, the fact that a weapon can conceivably be used in conformity with all applicable standards in a particular case does not mean that it should be allowed. The risk of misuse may be so high that it calls for a general prohibition.

Consider the following example: a high-value military commander may be inside a building, with computers containing information on his organisation's plans and structures. A number of uninvolved civilians also happen to be in the building. Depending on how high the value of the target is, and how many civilians there, it may in some cases be acceptable under IHL to bomb the building from the sky and thus to flatten it. However, if a weapon is available which can enter the building and identify and attack only the person in question, the civilians as well as the information on the computers as well may be spared. Depending on the specifics, this could in my view be a case where meaningful human control could potentially be exercised *inter alia* because the specific target is identified by a human being; the weapons are deployed within a confined space and is programmed to strike soon after deployment.

However, the question 'what are the chances of abuse' also needs to be asked before such machines are allowed to be made available on a wider scale. If it turns out that the danger of misuse of the weapon in situations where the human control is not as clear as in this example is significant, it may be necessary to forfeit the advantages of having such a system in isolated cases because of the larger scale implications. It is, for example, possible to conceive of a specific case where anti-personnel landmine may otherwise be used in conformity with international standards, but the risk of abuse is so high that it should be banned.

Simply relying on an application of the existing rules of IHL to individual cases to keep exponentially growing technology in check does not seem to allow for the growing tendency that human beings have to defer to technology when complicated decisions have to be taken. While the application of the rules of IHL to each such case in isolation may not point to the impermissibility of autonomous weapons, in aggregate the underlying values may be compromised to such an extent and the risk so compounded that a general normative intervention is called for. It is no exaggeration to say we are now at the point where we need to decide whether we want to retain meaningful control over targeting and indeed war itself, now and in the centuries to come.

## **5. Conclusions: what if fully autonomous weapons can save lives?**

If the use of machines, whether under meaningful human control or not, result in wrongful targeting, they should clearly not be used. In the case of those machines under meaningful human control, which can do proper targeting, the concerns raised above about the right to life and the right to dignity do not arise, and provided the general rules of IHL are met, there can be no objection to their use. There may indeed be an obligation to use them, should they be available.

This leaves the case of machines without meaningful human control, but they can ensure proper targeting. Let us assume for a moment that the answer to the 'can they

do it' question is positive in respect of a fully autonomous weapon. If that point is reached it will not take long before machine targeting will surpass human targeting, so we might as well focus on that scenario.

Assuming that such a scenario is reached and computers clearly surpass humans in many targeting situations (and many hold that it will be the case only under limited circumstances), it is clear that a prohibition on full autonomy potentially could come at the cost of human lives – lives that could otherwise have been saved. Autonomous weapons can thus become an effective violence reduction strategy – and indeed, those who do risk assessments with advanced software will probably be able to show one that very large numbers of lives may ultimately be at stake.

It is in this context necessary to consider the full implications of the approach suggested above, namely that even if robots can do proper targeting, fully autonomous targeting should not be permitted. Those who reach this conclusion based on human rights considerations may face a dilemma. Are they really in the name of human rights – and in particular, in the pursuit of vague considerations such as dignity – willing to forfeit technology that can save human lives? It could be argued that autonomous weapons are an armed conflict manifestation of the human rights imperative to use less lethal weapons in law enforcement and its use should be supported and indeed required from a human rights point of view. Calling for a ban on fully autonomous weapons, so the argument can run, flies in the face of human rights values.

There are several objections to this approach. It should in the first place be remembered that what is at stake is not a choice between a fallible and a perfect targeting system – neither humans nor machines will be infallible. We are thus required to compare the risks presented by the two alternatives, based on a judgement call about all the costs and benefits involved, including possible higher levels of safety in targeting, but also other concerns such as our notions of what a dignified life entails.

We certainly do not in other contexts always follow the route of always taking the safest option, irrespective of the other consequences. It remains a central tenet of the human rights project that torture is wrong on principled grounds and should be prohibited, even if in a particular case it could be argued – or proven – that it can save lives. The point of the ticking bomb scenario is to illustrate this very point. Likewise, even if mass surveillance can save lives in particular cases, for example, by laying terrorist plots bare, the human rights approach is that there is a point beyond which the value we place on privacy will not allow it, even if it will increase safety, because it challenges a central value of our society. The same reasoning, I would contend, applies to machines having the right of life and death over humans. Quality of life, the preservation of dignified life, and not mere continuation of biological life, is at stake.

Moreover, it follows from the discussion of the 'can they and should they' questions posed above that fully autonomous weapons do not entail a simple conflict between the right to life of civilians versus the dignity of those who may be targeted. Fully autonomous weapons present a potential threat to a range of rights, to different right-holders.

It is clear that autonomous weapons present a potential threat to the right to life of those who are protected by IHL from targeting, in that there will still be errors,



and there may be an accountability vacuum where these errors have to be addressed. As we saw, a lack of accountability is a violation of the right to life. It was also argued that autonomous targeting is inherently arbitrary, because it does not entail human deliberation, and thus violates the right to life even of those who may otherwise be targeted. In addition, autonomous targeting may also violate human dignity of a range of parties (those at the receiving end of force, but also those in whose name it is used). Even if it is possible that autonomous weapons may under certain circumstances spare lives (and to what extent this is the case in practice is contested), there are strong right to life and right to dignity arguments against the availability and use of such weapons if they have full autonomy.

If autonomy indeed means that weapons used in armed conflict become less lethal, and meaningful control is not relinquished, they should indeed be welcomed. However, if weapons are no longer tools in the hands of humans, human autonomy at the highest level is lost.

In the Western tradition, the middle ages saw religion as the central point of reference and justification of power. Seen from a secular perspective power was anonymous. This was replaced by a turn to language during the enlightenment, and the notions of the rule of law and the rise of human rights. Exercises of power had to be justified by providing reasons and deviations had to lead to accountability. This is the context in which the human rights project arose. The current turn to algorithms may in some respects be a step forward but it also presents the risk of a return to the anonymous exercise of power which does not have to meet the requirements of justification and accountability. There is a danger of anonymous force being exercised through autonomous machines. This is a challenge to its very roots that the human rights project has to face squarely.<sup>92</sup>

Human rights are not about numbers; ultimately it is about the quality and value of each individual life and of human life as such. Technology as a tool can greatly enhance the quality and value of human life; in fact, human life as we know it, including many of freedoms and rights, cannot exist on our planet without advanced technology. However, as the above case study of the use of technology in the extreme case of lethal weapons demonstrates, there are limits beyond which our transfer of power to technology should not go – not on the battlefield, and not in other realms of life. The weapon cannot be allowed to become the warrior.

The notion of ‘meaningful human control’ should be developed as a guiding principle not only for the use of autonomous weapons, but for the use of artificial intelligence in general; and not merely focussing on isolated uses of such technologies but focussing on the role of technology as such in our future. Allowing technology not only to supplement but indeed to replace human decision-making will undermine the very reason why life is valuable in the first place.

---

<sup>92</sup> I thank Markku Suksi from Abo Academy University in Finland for pointing me in this direction. See also M Suksi ‘On the Openness of the Digital Society: From Religion via Language to Algorithm as the Basis for the Exercise of Public Powers’ in A Lind, J Reichel, I Österdahl (eds), *Transparency in the Future – Swedish Openness 250 Years* (2017) (forthcoming).

## **Acknowledgements**

The article further develops ideas included in a report submitted by the author to the Human Rights Council (A/HRC/23/47) on 30 May 2013; in C Heyns 'Autonomous Weapons Systems: Living a Dignified Life and Dying a Dignified Death' in N Bhuta and others (eds) *Autonomous Weapons Systems: Law, Ethics, Policy* (2016) 3; C Heyns 'Human Rights and the Use of Increasingly Autonomous Weapons Systems (AWS) During Domestic Law Enforcement' (2016) 38 *Human Rights Quarterly* 350–78; as well as in various submissions made to meetings of the State Parties of the Convention on Conventional Weapons.

I thank Thompson Chengeta for his research assistance and Stu Maslen for valuable critical comments

## **Disclosure statement**

The author reports no conflicts of interest. The author alone is responsible for the content and writing of this article.

## **Notes on contributor**

*Christof Heyns* is a Professor of Human Rights Law at the University of Pretoria. He was the United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions 2010–2016.