

A stylized world map composed of a grid of dots in various shades of gray, with several dots highlighted in red. The title is overlaid on this map.

# The International-Law Dimension of Autonomous Weapons Systems

**ROBIN GEISS**  
October 2015

- Autonomous weapons systems are set to revolutionise how wars are fought. As things stand, there are still no completely autonomous weapons systems. But a trend towards more and more autonomy in military (weapons) systems is clearly discernible. This development should be subject to international regulation and control.
- The point of departure for future approaches to international regulation of autonomous weapons systems should be the basic assumption that »critical decisions« – that is, decisions that concern important legal interests such as, in particular, the right to life and the right to physical integrity – may not, for legal (international law) and ethical reasons, be delegated to completely autonomous systems. In particular, decisions concerning life and death must always be subject to the final decision and control of a human being.
- Against this background, Germany should seek to ensure, in international bodies such as the United Nations – especially within the framework of the UN Weapons Convention – but also beyond them, if necessary, that in the case of »critical decisions« the highest possible degree of human control is required and that completely autonomous lethal weapons systems are banned by regulations binding under international law.





<b>I. Introduction</b> .....	<b>3</b>
<b>II. Autonomous weapons systems: context and definition</b> .....	<b>3</b>
<b>III. Autonomous weapons systems: Present and future</b> .....	<b>6</b>
1. Definition .....	6
2. The state of the art .....	8
<b>IV. International Law and new weapons technologies</b> .....	<b>9</b>
<b>V. Legal and ethical issues</b> .....	<b>12</b>
1. Is the danger of armed conflicts increasing due to the deployment of autonomous weapons systems? .....	12
2. Can autonomous weapons systems comply with the rules of International Humanitarian Law? .....	13
3. Do Lethal Autonomous Systems Violate Human Dignity? .....	17
4. Is There an Obligation to Design Autonomous Weapons Systems to Act in a Non-lethal Manner? .....	18
5. The Problem of Liability .....	20
<b>VI. Proposals for Regulating Autonomous Weapons Systems</b> .....	<b>23</b>
1. Banning Autonomous Weapons Systems .....	23
2. Moratorium .....	24
3. Requirement of »Meaningful Human Control« .....	24
4. Curbing by Programming: The »Ethical Governor« .....	25
5. Approach Taken by the US Defense Department .....	25
6. Making Existing Regulations More Specific with Regard to Autonomous Weapons Systems .....	25
7. Laying Down New Rules and Higher Standards of Protection .....	26
8. Further Proposals .....	26
<b>VII. Conclusions and Recommendations for Action</b> .....	<b>27</b>



## I. Introduction

Arising from the controversy concerning the procurement of combat-capable drones for the German army the problem of fully automated or autonomous weapons technology is gradually coming to public attention. Predominant in this is a sense of unease, which is reflected in the positions of the political parties. While conflicting opinions are clearly evident in relation to drones,<sup>1</sup> there has so far been broad agreement in Germany about weapons that can operate autonomously, i.e. entirely without human control. Thus the coalition agreement between the CDU, the CSU and the SPD explicitly lays down that the Federal Republic of Germany »[will] advocate a ban under international law on fully automated weapons systems that deprive human beings of decision-making power over their deployment.«<sup>2</sup> In her speech to the Bundestag on 2 July 2014 German minister of defence Ursula von der Leyen reaffirmed Germany's position. The German army's plans on the procurement of combat drones concern only such unmanned systems with regard to which a human being always decides whether the weapon will be launched or not.<sup>3</sup>

In public debate the problems concerning the deployment of armed drones and of (future) fully autonomous weapons systems are often confused. However, it is important to analyse these two dimensions of modern weapons technology separately. The current debate on drones concerns remotely controlled (unmanned) aviation systems, in relation to which all processes are subject to human control and all critical decisions – for example, concerning target selection and weapon deployment – are always taken by military personnel. »Fully automated« or »autonomous« weapons systems, however, go a decisive step beyond this scenario. In

the future, these systems are supposed to be able to take over the selection and pursuit of targets and also the decision on firing the weapon independently. It is eminently possible to incorporate autonomous systems in drones. But the technology is by no means limited to that. Research is also being carried out on systems that could operate autonomously on the high seas, on land and in outer space. And while drone technology as such does not present any problems under international law – a drone is a weapon-carrying system comparable to a combat aircraft and only the mode of its deployment, especially with regard to targeted killings in an ostensible global war against terrorism is problematic in terms of international law – in the case of autonomous weapons systems, with regard to which life or death decisions are delegated to machines, it is the technology itself that gives rise to fundamental ethical and legal (international law) issues. These issues are the focus of the present study.

## II. Autonomous weapons systems: context and definition

Conventional drone technology has already changed the traditional understanding of the conduct of war. The remote control of drones makes deployment possible without direct personal risk from a distant control centre, on a computer screen. Experts, however, regard the development of autonomous systems as genuinely revolutionary and as a veritable paradigm shift in military technology.<sup>4</sup> This is because, in contrast to the current state of affairs, autonomous weapons systems will not only enhance the technical capabilities of the systems deployed. On top of that, in future only the basic decision to deploy the weapon will fall to a human being. All ensuing decisions concerning specific deployment on the ground would be handed over to the autonomous system.<sup>5</sup>

Currently, there are no genuinely autonomous weapons systems. Some robot engineers even doubt whether

1. See, for example, »Es geht nicht um automatisierte Killerdrohnen« [It's not about automated killer drones], *welt.de*, 30 June 2014, <http://www.welt.de/politik/deutschland/article129635441/Es-geht-nicht-um-automatisierte-Killerdrohnen.html> (accessed on 10 February 2015); »SPD skeptisch gegenüber Kampfdrohnen« [SPD sceptical about combat drones], *tagesspiegel.de*, 30 June 2014, <http://www.tagesspiegel.de/politik/opposition-warnt-vor-enthemmung-spd-skeptisch-gegenueber-kampfdrohnen/10127944.html> (accessed on 10 February 2015).

2. Deutschlands Zukunft gestalten [Shaping Germany's future], coalition agreement between the CDU, the CSU and the SPD, 18th Legislative period, 27 November 2013, p. 124, <http://www.bundesregierung.de/Content/DE/StatistischeSeiten/Breg/koalitionsvertrag-inhaltsverzeichnis.html> (accessed on 10 February 2015).

3. Speech by Germany's minister of defence Ursula von der Leyen to the Bundestag, 2 July 2014, <http://www.bundestag.de/dokumente/textarchiv/2014/kw27-de-drohnen/286034> (accessed on 10 February 2015).

4. For example, Peter Singer, *Wired for War*, 2009, pp. 179 ff.

5. UN General Assembly, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, A/HRC/23/47, 9 April 2013, marginal number 28, <http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47-en.pdf> (accessed on 10 February 2015).

such systems could ever be developed.<sup>6</sup> Most technical experts, however, assume that it is only a matter of time before such systems are operational. The US Department of Defense has drawn up an official plan to develop and bring into service autonomous systems increasingly up to 2038.<sup>7</sup> At the same time, the International Committee of the Red Cross (ICRC), in a speech before the United Nations, rightly pointed out that even today more and more »critical functions« in weapons systems operate autonomously, i.e. without human intervention.<sup>8</sup>

Based on such announcements many experts already expect a resumption of the international arms race.<sup>9</sup> The exigencies of military technology seem to make this inevitable: already today it is a problem for the armed forces of technologically highly advanced countries that the mass of information and the requirement of rapid decision-making and response in deployment are increasing constantly and that it is ever more difficult for human decision-makers to handle the amount of information, especially under combat conditions. To many observers it is therefore almost inevitable to press ahead with the development of increasingly autonomous systems, further accelerating the arms race.<sup>10</sup>

The announcements by some states – above all the United States – that in the future they will rely increasingly on autonomous military technology have given rise to an international debate on the ethical and legal implications of such systems in recent years. To date, the debate has focused primarily on »combat robots« (»Lethal Autonomous Robots« or LARS and »Lethal Autonomous Weapons Systems« or LAWS); in

other words, autonomous decisions at the end of the military decision-making chain in the context of selecting and attacking specific targets. The present study follows this (fairly narrow) focus, while also taking note of the fact that combat robots are most likely merely the tip of the iceberg and only one aspect of the challenges that go hand in hand with increasing autonomisation. After all, autonomous systems will gradually become increasingly important at all (even higher) levels of military and strategic decision-making.

In the eyes of their advocates, autonomous weapons systems have many advantages. For example, they are much more capable than human beings in gathering and processing information. They can operate more precisely, rapidly and flexibly. Autonomous systems would replace human beings on the battlefield and thus directly reduce the danger of human losses in armed conflict. Furthermore, they would not be influenced by emotions or physical exhaustion and thus be more capable than human soldiers of performing tiring routine tasks or very dangerous missions. Finally, because of their lack of fear, anger or hatred in stressful situations there would be no risk of excess.<sup>11</sup>

The critics of this technology, by contrast, warn that human life would be devalued if life and death decisions were ceded to a machine. The absence of emotion also has a reverse side, namely that autonomous weapons systems acting without emotion also lack empathy. Apart from that, the »video game mentality« observable in the case of conventional drone deployments on the part of the soldiers firing the weapons could be further reinforced if human beings only take the abstract decision whether to deploy autonomous systems or not.<sup>12</sup> It is also far from clear whether such systems could indeed be programmed in such a way that the danger of serious malfunctions could be (adequately) ruled out with certainty. Another view sometimes put forward is that a kind of weapon that reduces the acute risks of the party to the conflict deploying it to zero is inherently unethical, because of the significant asymmetry it would cause.

6. See Mary Ellen O'Connell, »Banning Autonomous Killing: The Legal and Ethical Requirement that Humans Make Near-Time Lethal Decisions«, in Matthew Evangelista and Henry Shue (eds), *The American Way of Bombing. Changing Ethical and Legal Norms, from B-17s to Drones*, Ithaca 2014, pp. 224, 226.

7. US Department of Defense, *Unmanned Systems Integrated Roadmap FY2013–2038*, 2013, <http://www.defense.gov/pubs/DOD-USRM-2013.pdf> (accessed on 10 February 2015).

8. United Nations, General Assembly, 69th session, First Committee, statement by the ICRC, New York, 14 October 2014, <https://www.icrc.org/en/document/weapons-icrc-statement-united-nations-2014> (accessed on 11 February 2015).

9. See, for example, the head of the »Technology and Ethics Study Group« at the Interdisciplinary Center for Bioethics at Yale University, Wendell Wallach, »Terminating the Terminator: What to Do about Autonomous Weapons«, *Science Progress*, 29 January 2013, <http://scienceprogress.org/2013/01/terminating-the-terminator-what-to-do-about-autonomous-weapons/> (accessed on 10 February 2015).

10. Hans-Arthur Marsiske, »Können Roboter den Krieg humanisieren?« [Can robots humanise war?], *Telepolis*, 13 April 2014, <http://www.heise.de/tp/artikel/41/41439/1.html> (accessed on 10 February 2015).

11. Geneva Academy of International Humanitarian Law and Human Rights, *Academy Briefing No. 8: Autonomous Weapon Systems under International Law*, November 2014, p. 4, <http://www.geneva-academy.ch/academy-publications/academy-briefings/1190-briefing-no-8-autonomous-weapons-systems-under-international-law> (accessed on 10 February 2015).

12. *Ibid.*, p. 5.

In particular, NGOs in the area of human rights and international humanitarian law have joined those voicing criticism. Thus, the International Committee for Robot Arms Control was founded as early as 2009.<sup>13</sup> In October 2012 several NGOs joined forces in the Campaign to Stop Killer Robots,<sup>14</sup> in order to advance the debate on autonomous weapons systems at the international level. The European Parliament, too, has in the meantime called for an outright ban on the development, production and use of completely autonomous weapons systems.<sup>15</sup>

The scientific and political debate on the benefits and dangers of autonomous weapons is embedded in a broader social debate on the implications of the increasing tendency towards automation in many areas of life. The military dimension of this debate is only the tip of the iceberg. Because the absolutely fundamental question arises of how much »dehumanisation« global society can (and wants to) bear in its everyday functioning before the social costs become too high or in any case incalculable.<sup>16</sup> The problems that arise, for example, in critical situations due to the increasing automation of civil aviation have often been remarked upon.<sup>17</sup> Besides purely technical issues concerning the safety of such systems it is increasingly the ethical dimension that determines the public debate. If computer guided machines perform tasks independently in more and more areas of life, society must clarify for itself how the algorithms should allow the controlling computer to »decide« in borderline moral situations. Where should a fully automated car, for example, swerve if a child is running in the street?<sup>18</sup> Should medical robots be able to prescribe strong painkillers to patients independently, without a doctor monitoring

decisions in individual cases?<sup>19</sup> Obviously, the possibility, on top of all this, that computer-controlled systems could make autonomous decisions on the offensive killing of people gives rise to considerably more unease.

Technological development in these areas often appears to be inevitable, as a result of which it is no longer regarded as beneficial by a large part of society. Ethics and (international) law thus serve as »fettering systems« for the purpose of banning technology that has become a threat.<sup>20</sup> This applies just as much at the national as at the international level.

Against this background, a broad public debate is urgently needed on autonomous weapons systems. Even though in 2015 we are still a long way away from seeing human armies replaced by robots, society already has to consider the ethical, political and legal implications of such systems. The politics of international law is generally inclined to tackle the consequences of a new technology only after it has already come into use. Traditionally, international law has typically sought to regulate new weapons technology at least one war too late. Such an approach – especially when it concerns not just a single weapon but a paradigm change in military technology – potentially has far-reaching consequences. Even though, technically, much of this is still speculation, we must make a start on clarifying the issues. Occasionally, critics even claim that any debate would already be too late. So many systems in combat use are already automated that, due to the ensuing information inundation, human soldiers are simply no longer in a position to make the relevant decisions unaided.<sup>21</sup> In other words, the dehumanisation (in the literal sense) of war has already begun. We are only a short step away from fully autonomous weapons systems.

13. See: <http://icrac.net/>.

14. See: <http://www.stopkillerrobots.org/>.

15. Resolution of the European Parliament on the deployment of armed drones, 2014/2567(RSP), 25 February 2014, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+MOTION+P7-RC-2014-0201+0+DOC+XML+V0//DE> (accessed on 10 February 2015).

16. Paul Ford, »Our Fear of Artificial Intelligence«, *MIT Technology Review*, 11 February 2015, <http://www.technologyreview.com/review/534871/our-fear-of-artificial-intelligence/> (accessed on 16 February 2015).

17. See, for example, William Langewiesche's analysis of the crash of Air France flight 447, »The Human Factor«, *Vanity Fair*, October 2014, <http://www.vanityfair.com/news/business/2014/10/air-france-flight-447-crash> (accessed on 11 February 2015); Nicholas Carr, *The Glass Cage: Automation and Us*, New York 2014.

18. Zetsche calls for an ethical debate on driverless cars, *Frankfurter Allgemeine Zeitung*, 8 January 2015, <http://www.faz.net/aktuell/wirtschaft/unternehmen/daimler-chef-dieter-zetsche-fordert-ethische-debatte-ueber-roboterautos-13358455.html> (accessed on 11 February 2015).

19. Robin Marantz Henig, »Death by Robot«, *The New York Times Magazine*, 9 January 2015, [http://www.nytimes.com/2015/01/11/magazine/death-by-robot.html?\\_r=0](http://www.nytimes.com/2015/01/11/magazine/death-by-robot.html?_r=0) (accessed on 11 February 2015).

20. Herfried Münkler, Neue Kampfsysteme und die Ethik des Krieges [New combat systems and the ethics of war], speech at the Heinrich-Böll-Stiftung, 21 June 2013, <http://www.boell.de/de/node/277436> (accessed on 11 February 2015).

21. Niklas Schörnig, Automatisierte Kriegsführung – Wie viel Entscheidungsraum bleibt dem Menschen? [Automated warfare – how much decision-making room remains for human beings?], *Aus Politik und Zeitgeschichte* 35–37/2014, 18 August 2014, <http://www.bpb.de/apuz/190115/automatisierte-kriegsfuehrung-wie-viel-entscheidungsraum-bleibt-dem-menschen?p=all> (accessed on 11 February 2015).

Against the background just outlined the present study will examine the international-law and ethical implications and limits of the deployment of autonomous weapons systems. The analysis begins with a presentation of the technological »state of the art« and defines the notion of »autonomous weapons system« (Section 3). Then, by way of an overview, we take a historical look at international law's treatment of new weapons technologies (Section 4), before breaking down and analysing the emerging ethical and (international) legal issues related to autonomous weapons systems (Section 5). After presenting current proposals for banning autonomous systems (Section 6) the study concludes with specific recommendations for action (Section 7).

### III. Autonomous weapons systems: Present and future

#### 1. Definition

In the scientific literature and official government documents there are a number of approaches to defining autonomous weapons systems. At present, there is no universally accepted definition. Common to all the different approaches, however, is that the level of capability with regard to decision-making by means of algorithms alone, without human intervention, is stressed as the decisive criterion for »autonomy«. Autonomy should therefore not be understood in the moral philosophical sense as the free will of an individual. Even an autonomous robot can operate only within the limits of the possibilities programmed into it by means of algorithms. However, to conclude from this that no really autonomous system can exist, because at a certain point a human being is always involved in the decision-making matrix,<sup>22</sup> is an unjustified simplification of the problems arising. Those who pursue this line of argument have no problem with autonomous systems. In their view ultimately, everything remains as it was. Every decision made by the supposedly autonomous system would be traceable to a human decision. This standpoint, however, fails to recognise the level of autonomy that military technology research is striving to reach in the long term. Needless to say, human

influence is never completely excluded. People decide whether a deployment should take place and, in the course of programming, its framework conditions. And it is – for the time being – human decision-makers who activate the robots and then deactivate them again. But when the decision on at least a potentially lethal use of weapons is taken significantly before actual deployment one can no longer talk of any genuine human control at the key moment. In highly complex conflict situations in particular the (unforeseen) scenarios that the despatched systems will be confronted with in the course of their missions are no longer predictable.

In contrast to autonomous systems automated systems merely automatically execute previously programmed commands in a predictable fashion. They are unable to react independently to unforeseen events. Such systems are typically deployed to carry out an entirely determinate, pre-established task (for example, defence against anti-ship missiles) in a narrowly defined area of operations (for example, 5 nautical miles around the defending ship) and time frame. A mine also falls under – as the simplest possible example – this conceptual definition of automated weapons. The more complex the field of functions and the larger or more differentiated the area of deployment, the more the distinction between autonomous and automated systems becomes blurred. Where exactly the line is to be drawn along this continuum cannot be said with any certainty. A clear-cut distinction between autonomous and automated systems is not (always) possible. However, it is probably not even required for clarification of the relevant ethical and legal problems. The International Committee of the Red Cross (ICRC) has now largely abandoned any distinction between autonomous and automated systems. According to the ICRC, what is decisive instead is whether critical functions – that is, independent decision-making about life and death – are entrusted to a system.<sup>23</sup> This approach is a convincing one. Regardless of the technological definitions, what is decisive from an ethical and an (international) legal standpoint is where the line is to be drawn between delegatable decisions and decisions that must remain subject to direct human control. The ICRC's approach also makes it clear that not every autonomous system is necessarily problematic.

22. Thus, to some extent, Michael N. Schmitt and Jeffrey S. Thurnher, »Out of the Loop«: Autonomous Weapons Systems and the Law of Armed Conflict, *Harvard National Security Journal*, Vol. 4, 2013, pp. 231, 280, although they certainly recognise that problems arise from the deployment of such systems.

23. United Nations, General Assembly, 69th session, First Committee, statement by the ICRC, New York, 14 October 2014, <https://www.icrc.org/en/document/weapons-icrc-statement-United-nations-2014> (accessed on 11 February 2015).

An autonomously operating submarine used to search for mines or an autonomous offensive system deployed exclusively in outer space to neutralise hostile military satellites do not raise the same kind of urgent ethical and legal issues as a combat robot designed for urban warfare that directly encounters human beings, both enemy combatants/fighters and civilians.

Regardless of the existing delimitation problems, in what follows the current models and definitional approaches commonly used in the identification and categorisation of autonomous systems are presented because the international debate so far has essentially relied on these (step based) models.

UN Special Rapporteur Christof Heyns describes autonomous weapons systems as robots that gather information about their surroundings by means of sensors, which is then processed so that a decision can be taken, which is finally executed by the installed components (for example, weapons or means of transport).<sup>24</sup> In accordance with the (broad) definition of the US Department of Defense, weapons systems are autonomous if, after their activation, they independently – that is, without further human intervention – select targets and attack. There is also autonomy when human decision-makers are still able to abort specific decisions to attack even after activation.<sup>25</sup> The International Committee of the Red Cross, in turn, also considers decisive the independence of the installed critical systems with regard to target recognition, tracking, selection and attack.<sup>26</sup>

The NGO Human Rights Watch, in its study on combat robots, takes a three-step approach, which has since become widely adopted in the literature.<sup>27</sup> The definition's starting point is the level of human participation in the actions of the system with reference to the individual functions of observation, orientation, decision-making and action in response to a given situation or a certain

environment.<sup>28</sup> On this basis Human Rights Watch differentiates whether, with regard to a weapons system, a human being is »in the loop«, »on the loop« or »out of the loop«. »In the loop« means that the machine cannot operate without human decision-making. One or several decisive steps lie in the hands of the guiding person. Such »human in the loop« systems are not autonomous. The drones currently being deployed by the United States over Pakistan, Afghanistan and Yemen, which are controlled remotely by a distant command centre, belong in this category. So-called semi-autonomous systems also fall into this category. This includes weapons whose target selection is carried out by human beings, but which then independently execute target tracking and destruction. According to the definition these systems are still not to be classed as autonomous. »On the loop« means that the machine is basically in a position to execute all steps independently, without human intervention. Human beings monitor the system's actions and can intervene if necessary in order to abort the machine's decisions. This variant can also be designated monitored autonomy. In his study, however, UN Special Rapporteur Heyns rightly points out that, in reality, the intervention option can be extremely limited if a robot takes decisions in a matter of milliseconds. In this instance meaningful control is de facto no longer possible.<sup>29</sup>

If human beings, finally, are »out of the loop« the system is operating completely autonomously. There are no longer any direct opportunities for intervention. The machine operates independently with regard to all necessary steps in carrying out an attack. Furthermore, it is expected that future generations of robots will have the capability to learn from »experience« and thus to adapt their functioning independently and perhaps even extend it.<sup>30</sup>

Professor of artificial intelligence and robotics Noel Sharkey, by contrast, proposes a level-based model with two added categories.<sup>31</sup> In his view a total of five levels of human control over weapons systems should be distinguished: (i) the human deliberates about a target

24. Heyns, see above, note 5, marginal number 38.

25. US Department of Defense, *Autonomy in Weapons Systems*, Directive No. 3000.09, 21 November 2009, pp. 13f, <http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf> (accessed on 11 February 2015).

26. Report of the Expert Meeting on Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects, 26–28 March 2014, Geneva, 9 May 2014, p. 1, <https://www.icrc.org/eng/assets/files/2014/expert-meeting-autonomous-weapons-icrc-report-2014-05-09.pdf> (accessed on 11 February 2015).

27. Human Rights Watch, *Losing Humanity: The Case against Killer Robots*, 2012, p. 2, <http://www.hrw.org/reports/2012/11/19/losing-humanity-0> (accessed on 11 February 2015).

28. On this see William Marra and Sonia McNeil, *Understanding »The Loop«: Regulating the Next Generation of War Machines*, *Harvard Journal of Law and Public Policy*, Vol. 36, 2013, p. 1139.

29. Heyns, see above note 6, marginal number 41.

30. Schmitt/Thurnher, see above note 22, p. 240.

31. Noel Sharkey, *Towards a Principle for the Human Supervisory Control of Robot Weapons*, 2014, p. 11, <https://www.mini-symposium-tokyo.info/ICRA2014/sharkey2014.pdf> (accessed on 11 February 2015).



before initiating any attack; (ii) a computer program provides a list of targets and a human being chooses which one to attack; (iii) the program selects the target and a human being has to give his or her approval before an attack is carried out; (iv) the program selects the target and a human being has a restricted time in which to veto an attack; and (v) the program selects the target and initiates an attack without human involvement.

Both the threefold division of Human Rights Watch and Sharkey's five-level model are necessarily schematic and simplified. In reality, transitions are often blurred and not clear-cut.<sup>32</sup> Scenarios are imaginable and even probable in which human soldiers remain »in the loop« only formally, but important aspects of decision-making processes are delegated to a machine, such as target selection in accordance with Sharkey's level (iii). Even if decision-making on actual deployment of the weapon remains in human hands, one must assume that in stressful situations and under time pressure – typical features of any combat mission – the person involved will, in case of doubt, defer to the machine. Such behaviour – in other words, the tendency to trust in an automated machine even if there are substantial indications that it is unreliable or, in some cases, makes mistakes – is known as »automation bias«.<sup>33</sup> In such a case, however, human control – although technically possible – in reality is meaningless..

## 2. The state of the art

There are still no autonomous weapons systems in existence. As already mentioned, scientists are divided over when the development process will be concluded and the first systems fully operational. However, the autonomisation of weapons systems – that is, in defensive systems, which typically operate on the basis of a narrowly defined and predetermined deployment scenario – is proceeding apace. In particular with regard to target selection and the decision to launch an attack automation and increasingly also autonomisation tendencies are already clearly discernible. For purposes of illustration some of the systems that are already in

place today or that are currently in the final stages of development are listed below:

**Phalanx CIWS:** US close-in weapon system (CIWS) for defence against incoming missiles, such as anti-ship missiles. It can recognise and attack targets autonomously.<sup>34</sup>

**Iron Dome:** Israeli mobile defence system to defend against short-range missiles. Like Phalanx, it is used only for defensive purposes. The system comprises several components, which interact fully automatically. If the radar recognises the launch of a hostile missile, its trajectory is calculated and the information passed on to the control centre, which determines the point of impact. If this point lies within one of the previously determined civilian defence zones, a soldier in the field manually confirms the launch of the interceptor missile. The last step is thus executed by a human being.<sup>35</sup>

**BAE Taranis:** British prototype of an autonomous combat drone that seeks and identifies targets autonomously. It is supposed to be able to attack only if a direct order is given by a human operator.<sup>36</sup>

**US Northrop Grumman X-47B:** US drone prototype, the first autonomous system to land independently on an aircraft carrier. Airborne navigation is also supposed to be possible autonomously.<sup>37</sup>

**Samsung SGR-A1:** South Korean robot, which guards the demilitarised zone at the border with North Korea. The system is armed and is supposed to be able to recognise and attack enemy combatants independently.<sup>38</sup>

32. Schönrig, see above note 21.

33. Peter M. Asaro, Modelling the Moral User, *IEEE Technology and Society Magazine*, 2009, p. 22, <http://peterasaro.org/writing/Asaro%20Modeling%20Moral%20User.pdf> (accessed on 11 February 2015).

34. <http://de.wikipedia.org/wiki/Phalanx-CIWS> (accessed on 11 February 2015).

35. Meet Israel's home-front hero: Iron Dome, *Haaretz*, 18 July 2014, <http://www.haaretz.com/business/.premium-1.605770> (accessed on 11 February 2015).

36. See the manufacturer's product description: <http://www.baesystems.com/cs/groups/public/documents/document/mdaw/mtg4/-edisp/baes-165043.pdf> (accessed on 11 February 2015).

37. <http://en.wikipedia.org/wiki/Northrop-Grumman-X-47B> (accessed on 11 February 2015).

38. Robocop soll die innerkoreanische Grenze schützen [Robocop to defend internal Korean border], *heise.de*, 25 January 2007, <http://www.heise.de/newsticker/meldung/Robocop-soll-die-innerkoreanische-Grenze-schuetzen-138623.html> (accessed on 12 February 2015).

**Guardium:** Driverless car that autonomously patrols the border between Israel and the Gaza Strip. Up to now it has been unarmed, but according to the Israel Defence Forces it can be armed and then take action autonomously against enemy movements.<sup>39</sup>

**IAI Harpy:** Drone developed by Israel which independently locates and can attack enemy radar sites. It is also deployed by, among others, Turkey, China, South Korea, Chile and India.<sup>40</sup>

**Brimstone:** British so-called »fire and forget« missile, which can automatically identify tanks and other vehicles and is able, within a previously defined area, to find targets autonomously. After operational experience in Afghanistan the system was upgraded to enable, if need be, a human being to choose the targets if the missile is deployed in an area in which there are also civilians or one's own or allied troops.<sup>41</sup>

**GuardBot:** Amphibian robot designed to support human troops on beach landings, developed by a private US firm in cooperation with the US Navy. The system is currently still remotely controlled, but ultimately it is to be provided with more and more autonomous functions, so that it can independently head for previously determined points or can search potential landing zones for explosive devices.<sup>42</sup>

Besides these weapons technologies autonomous systems are also being developed in cyberspace. For example, the US National Security Agency (NSA) is working on a program called MonsterMind, which is supposed to be able to capture data streams independently in order to neutralise cyber attacks on the United States immediately and automatically.<sup>43</sup> It is not clear whether and when the system will become operational.

39. See the manufacturer's product description: <http://www.iai.co.il/Sip-Storage/FILES/0/33810.pdf> (accessed on 12 February 2015).

40. See the manufacturer's product description: <http://www.iai.co.il/Sip-Storage/FILES/4/34884.pdf> (accessed on 12 February 2015).

41. [http://en.wikipedia.org/wiki/Brimstone-\(missile\)](http://en.wikipedia.org/wiki/Brimstone-(missile)) (accessed on 12 February 2015).

42. Patrick Tucker, The Marines Are Building Robotic War Balls, *Defense One*, <http://www.defenseone.com/technology/2015/02/marines-are-building-robotic-war-balls/105258/> (accessed on 17 February 2015).

43. James Bamford, The Most Wanted Man in the World, *Wired*, August 2014, <http://www.wired.com/2014/08/edward-snowden/> (accessed on 12 February 2015).

## IV. International Law and new weapons technologies

Modern international law in the course of its development has time and again been confronted with new kinds of weapons systems, accompanied by constant and sometimes rapid technological development.<sup>44</sup> One hundred years ago the First World War ushered in an unprecedented mechanisation of war. Technical change is thus very much the norm. Nevertheless, it is always a challenge for international law, with its static and often protracted law-making processes. This applies in particular when what is at issue is not only a determinate new kind of weapon – a new artillery shell, for example – but a technology that, like increasing autonomisation, is capable, across the board, of structurally changing or even revolutionising military deployments and the conduct of war in general.

Autonomous weapons systems highlight a whole spectrum of partly familiar, but also entirely new issues and problems. Autonomous combat robots serve the purpose, among other things, of sparing human soldiers on one's own side. The risk of losses is minimised if soldiers no longer have to be present on the battlefield. On this reasoning autonomous weapons systems belong to the long history of developing so-called »ranged« weapons, beginning with the invention of the bow and arrow and continuing through the crossbow, gunpowder, artillery, the air force and now drones. The introduction of such technology, which was aimed at reducing the need for direct confrontation with the enemy, has always been criticised as unethical or even »unchivalrous«.<sup>45</sup> From this viewpoint at least – minimising risk – autonomous weapons systems are not fundamentally new and, in particular, closely linked to the debate on drones. Furthermore, the autonomisation of weapons technology also – and in particular – serves the purpose of mastering the flood of information and data characteristic of contemporary conflict situations and »optimising« military decision-making processes. In this way human decision-makers are potentially excluded from important and critical decisions in many areas and replaced by computer-based decision-making. This

44. Robin Geiß, The Law of Weaponry from 1914 to 2014. Is the Law Keeping Pace with Technological Evolution in the Military Domain?, in Jost Delbrück et al. (ed.), *Aus Kiel in die Welt: Kiel's Contribution to International Law*, Berlin 2014, pp. 229, 237.

45. Münkler, see above note 20.

change brings to light fundamentally new ethical and legal issues.

In modern international law the regulation and prohibition of weapons systems has always taken place on two different levels. On one hand, the prohibition of certain (kinds of) weapons was agreed on. The ban treaties concluded for this purpose are based on the humanitarian considerations of the laws of armed conflict, especially as anchored in the Geneva Conventions and their Additional Protocols, which is designated synonymously as international humanitarian law and as *ius in bello*. According to the basic idea underlying international humanitarian law certain kinds of weapons are inherently unethical, whether because they cause unnecessary suffering or whether because, by their very nature, they are indiscriminate – that is, they attack protected civilians and combatants alike and without distinction. On the other hand, there are the disarmament treaties, which emerged in particular after the Second World War and in the course of the Cold War. They are bound to the prohibition of war and the use of force in international relations enshrined in the UN Charter, the so-called *ius contra bellum*: a global reduction of the number of weapons is supposed to help to reduce the danger of any use of force between states. More recently it may be observed that the boundaries between weapons bans and disarmament treaties are becoming more blurred. Originally arising from mere prohibitions on use – for example, the Geneva Gas Protocol of 1925<sup>46</sup> – agreements today increasingly contain more comprehensive regulations, which also extend to the research, development, storage and sale of such systems. The Chemical Weapons Convention of 1993 is one such treaty, which combines elements of disarmament and humanitarian considerations and is no longer confined to prohibiting a specific weapon, but rather a whole category of weapons.<sup>47</sup>

A look into the past shows that weapons ban treaties usually come »one war late«. For example, states were able to reach agreement on the Geneva Gas Protocol in 1925 only after poison gas had been deployed in the First

World War, with devastating results. More recently, too, treaty bans on anti-personnel mines and cluster bombs were effected only after protracted negotiations and decades of humanitarian lobbying by NGOs. As far as can be seen, the only weapons that, despite having been developed to the point of operability, could be banned before they were actually deployed in armed conflict are blinding laser weapons. These weapons' *modus operandi* – which would have resulted in permanent blindness – was considered to cause cruel and unwarranted suffering. Having said that, there is little doubt that they could be banned so rapidly mainly because their military and strategic potential was considered to be fairly low. The case of cluster bombs is quite different. Their ban was based on their indiscriminate effects and devastating consequences for the civilian population. Because cluster bombs were considered to have major military potential, however, a number of militarily important countries – such as the United States, Russia and China – have not ratified the ban treaty.

Because it takes so long to conclude international law treaties and even once concluded a treaty is still reliant on attaining universal ratification, certain general principles were enshrined in the humanitarian legal framework that are intended to ensure the dynamics and fundamental adaptability of such legislation to technological change. The key »principle of distinction« (differentiation) in common with the fundamental prohibition of causing unnecessary suffering, applies to all weapons systems, whether old or new. This was confirmed by the International Court of Justice in its advisory opinion on nuclear weapons, which are not mentioned explicitly in the Geneva Conventions and their Additional Protocols.<sup>48</sup> These principles are abstract and therefore timeless. They apply regardless of the state of technology and can also be applied to entirely new kinds of technology, such as cyberweapons or autonomous weapons systems. At the same time, the high level of abstraction of these principles means that there is always room for discussion and so, in the long term, it is always worth pursuing a specific ban treaty in which all technical details of a particular (new) technology can be regulated. In the past, however, these fundamental principles have often provided the decisive impetus in getting a new weapons technology onto the agenda in treaty negotiations.

46. Protocol for the Prohibition of the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, <https://www.icrc.org/ihi/INTRO/280?OpenDocument> (accessed on 13 February 2015).

47. <http://www.auswaertiges-amt.de/DE/Aussenpolitik/Friedenspolitik/Abraestung/BioChemie/Verbot-C-Waffen-node.html> (accessed on 13 February 2015).

48. ICJ, Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons, ICJ Reports 1996 (I), recitals 84, 85.

Other important general clauses that, beyond the abovementioned principles, are supposed to ensure the timelessness and dynamism of international humanitarian law include the so-called Martens Clause and the weapons inspection obligation enshrined in Article 36 of the First Additional Protocol to the Geneva Conventions.

The Martens Clause, which was first put in writing in the preamble of the Hague Convention of 1899,<sup>49</sup> finds its contemporary formulation in Article 1(2) of the First Additional Protocol to the Geneva Conventions:<sup>50</sup>

*In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.*

The International Committee of the Red Cross concludes from this that new weapons must be investigated before their deployment to see whether they are at odds with the basic principles of humanity or the dictates of public conscience.<sup>51</sup>

Closely linked to the Martens Clause is Article 36 of the First Additional Protocol to the Geneva Conventions. It attaches conditions to the deployment of new weapons:

*In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.*

In other words, the agreement lays down the obligation of the states parties to the First Additional Protocol, before the introduction of new weapons technologies, to investigate whether they would violate the provisions of international humanitarian law or any other applicable regulation of international law. Although the customary

law status of the clause is controversial, even some militarily important states, which have not ratified the First Additional Protocol, have at least recognised the fundamental obligation to inspect weapons and have introduced formal investigation procedures.<sup>52</sup> However, it remains controversial whether a weapon that violates Article 36 Additional Protocol I is automatically prohibited or whether a special treaty is required for that. The more convincing view is that a weapon cannot be permitted if it is not compatible with the principles of international humanitarian law per se. A clear difficulty in this respect is that – as already mentioned – the principles themselves are not sufficiently precise and tend to apply rather high standards. A new weapon is therefore only rarely deemed to be in violation thereof.<sup>53</sup>

With regard to the development of autonomous weapons systems it has rightly been pointed out that it is crucial to commence their investigation early on, based on the abovementioned principles, and in particular to keep an eye on the process at every step. Because if the weapons are manufactured solely on the basis of purely technical feasibility and then are made ready for sale, given the immense research and development costs a vested interest arises that makes the commissioning of such systems virtually inevitable. In such circumstances and at that point it would seem difficult to arrive at the conclusion that there is no compatibility with Article 36 Additional Protocol I.<sup>54</sup>

Apart from that, another, more fundamental question arises in connection with the inspection mechanisms for new weapons systems. Although it is true that international humanitarian law, with its provisions for the legal inspection of new weapons systems to which we have referred, is sufficiently dynamic and can adapt itself to new developments, at the same time one can hardly escape the conclusion that the regulations are static at least to the extent that they are still based on the same ethical principles that were laid down over 100 years ago. The prohibition of unnecessary suffering

49. <http://www.admin.ch/opc/de/classified-compilation/18990009/index.html> (accessed on 13 February 2015).

50. <http://www.admin.ch/opc/de/classified-compilation/19770112/index.html> (accessed on 13 February 2015).

51. ICRC, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare*, Geneva 2006, p. 17, <https://www.icrc.org/eng/assets/files/other/icrc-002-0902.pdf> (accessed on 13 February 2015).

52. See Human Rights Watch, see above note 27, p. 22; this applies in particular to the United States, see US Department of Defense Directive 5000.01: The Defense Acquisition System, Defense Acquisition Guidebook, E1.1.15 (Legal Compliance), 12 May 2003, <https://acc.dau.mil/CommunityBrowser.aspx?id=314789> (accessed on 13 February 2015).

53. Geiß, see above, note 44.

54. Marco Sassòli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified*, *International Law Studies*, Vol. 90, 2014, pp. 308, 322.

and in broad outline also the principle of distinction existed in international law before the First World War. If one takes the view, however, that autonomous weapons systems are not merely an evolution of existing weapons technology, but are to be categorised as ushering in a new era, it must at least be put up for discussion whether the inherited principles of international humanitarian law – regardless of their undoubted continuing relevance – in and of themselves are still equipped to cope adequately with this entirely new kind of weapons systems.

In the next section we must therefore look in detail at the extent to which the fundamental principles of international humanitarian law can still meaningfully apply to autonomous weapons systems, as well as the extent to which these systems may give rise to ethical and legal questions and challenges beyond that.

## V. Legal and ethical issues

In what follows the most important legal and ethical issues concerning autonomous weapons systems are presented and analysed. First of all, it is important to recognise in this respect that autonomy in and of itself is not the problem. The introduction of independently operating mine hunters or robots that can be deployed for bomb disposal<sup>55</sup> can easily be justified. It is rather – as already mentioned – the delegation of critical decisions to non-human systems that gives rise to unease and throws up fundamental ethical and legal issues. Particularly critical is the delegation of decisions concerning human life. The examples of driverless vehicles and medical robots, mentioned above, show that the issue is not confined to military weapons systems. With regard to the latter the legal and ethical challenges come to the fore, however, because it is the very purpose of these systems to cause harm.

But even with regard to autonomous weapons systems by no means all scenarios encounter the same challenges. References to a future »clean« – that is, ethically and legally unobjectionable – war between robots alone, however, appear as illusory as the idea of purely virtual conflicts in cyberspace. Although it is true that

in a theoretical scenario in which only combat robots from each side encounter one another, human legal interests such as the right to life would not be (directly) affected.,<sup>56</sup> it has also rightly been remarked that it can hardly be reasonably predicted what would happen if the autonomous weapons systems of two hostile states encounter one another. The unpredictability of highly complex algorithms reacting to one another could lead to a scarcely controllable military escalation.<sup>57</sup> Above all, it cannot be assumed that the emphasis will be on such simulations. At least with regard to the increasing autonomisation of unmanned aircraft the opposite should be assumed, namely that in the first instance human legal interests will continue to be directly affected. It is precisely the asymmetrical conflict situations of the present that propel development in the direction of autonomous weapons systems.<sup>58</sup> And as long as armed conflicts involve achieving control over people and states by force of arms purely virtual scenarios will remain utopian.

### 1. Is the danger of armed conflicts increasing due to the deployment of autonomous weapons systems?

A critical point that is often raised in relation to autonomous weapons systems is that the threshold for the deployment of military force would be lowered by the technological development of such systems.<sup>59</sup> As things stand at the moment, predictions are difficult in this regard. As far as can be seen there are no robust empirical surveys in this area. In the first instance, then, we are dealing with an assumption, albeit perhaps an obvious one. In any event, the example of the much more intensive use of armed drones by the United States under President Obama demonstrates that weapons systems that drastically reduce the risk of losses on one's own side can even achieve public acceptance if the population of the country has grown »war weary« in the wake of past military deployments, as in Afghanistan and Iraq. If the overall costs to society of a military deployment

55. Hitherto such machines have been controlled remotely; see, for example, the telerob Explosive Ordnance Disposal and Observation Robot (tEODor), <http://www.army-technology.com/projects/teodor-explosive-ordnance-eod-robot/> (accessed on 14 February 2015).

56. For example, Peter M. Asaro, *How Just Could a Robot War Be?*, in Philip Brey et al. (eds), *Current Issues in Computing and Philosophy*, Amsterdam 2008, pp. 50, 62.

57. Schörnig, see above note 21.

58. Cf. Münkler; see above note 20: »Drones, robots and similar are instruments with which post-heroic societies will ward off asymmetrically acting heroic societies.«

59. Cf. Asaro, see above note 56.

diminish, even modern »post-heroic societies«<sup>60</sup> are in a position to get involved in military pacification projects.<sup>61</sup> In particular with regard to autonomous systems, that would appear to favour enhanced potential for armed deployments.<sup>62</sup>

Besides the risk minimisation aspect the significant psychological inhibition threshold with regard to killing is also removed or at least diminished if algorithms take over the decision-making. The deployment of weapons in this way appears increasingly clean and abstract from society's standpoint because the real consequences of armed engagement are scarcely perceptible. While this already applies – with some reservations – to targeted killings by means of drones, combat robots could further increase this effect. Weapons deployment would be normalised. And while post-traumatic stress disorder appears to be just as frequent among drone pilots as among regular combat units,<sup>63</sup> such consequences would disappear entirely with the deployment of robots. In that case, war would no longer constitute a state of emergency affecting society as a whole.<sup>64</sup> On the contrary, as the past decade of the »global war on terror« has shown, by means of the deployment of unmanned, autonomous weapons systems the elimination of the temporal and spatial boundaries of armed conflict could proceed even further.<sup>65</sup>

60. On this idea see Herfried Münkler, *Heroische und postheroische Gesellschaften*, Merkur 2007, p. 742.

61. Herfried Münkler, *Der Asymmetrische Krieg [Asymmetric war]*, *Der Spiegel* 44/2008, pp. 176, 177.

62. Philip Alston, *Lethal Robotic Technologies: The Implications for Human Rights and International Humanitarian Law*, *Journal of Law, Information & Science*, Vol. 21, 2011/2012, pp. 36, 55; Münkler, for example, however, expressly doubts that the use of drones is lowering the inhibition threshold for armed conflicts, justifying his view on the basis of the requirement for parliamentary approval; see interview with Herfried Münkler, »Drohnen zu ächten, wäre absurd« [It would be absurd to outlaw drones], *Stuttgarter Zeitung*, 17 July 2014, <http://www.stuttgarter-zeitung.de/inhalt.interview-mit-herfried-muenkler-drohnen-zu-aechten-waere-absurd.bed9bfd9-29bf-4b3c-810d-7aec7ad6299b.html> (accessed on 14 February 2015). However, this argument is applicable – if at all – only in the German context.

63. James Dao, *Drone pilots are found to get stress disorders much as those in combat do*, *New York Times*, 22 February 2013, [http://www.nytimes.com/2013/02/23/us/drone-pilots-found-to-get-stress-disorders-much-as-those-in-combat-do.html?\\_r=0](http://www.nytimes.com/2013/02/23/us/drone-pilots-found-to-get-stress-disorders-much-as-those-in-combat-do.html?_r=0) (accessed on 14 February 2015).

64. Peter M. Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making*, *International Review of the Red Cross*, Vol. 94, 2012, pp. 687, 692.

65. See also Heyns, see above note 5, recital 57 ff.; other experts point to the capability of autonomous weapons systems to spare the lives of both soldiers and civilians; consequently, any reference to an increase in military deployments would miss the point. See also Kenneth Anderson and Matthew C. Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, p. 18, Stanford University, The Hoover Institution, Jean Perkins Task Force on

## 2. Can autonomous weapons systems comply with the rules of International Humanitarian Law?

In the debate on autonomous weapons systems the focus to date has been on the question whether such systems would be capable of complying with the rules of international humanitarian law. These are the provisions of international law that are applicable during an armed conflict. They modify key human rights provisions, in particular the right to life. Thus, during armed conflict the right to life is granted only in accordance with international humanitarian law. That means, among other things, that enemy fighters and combatants may be attacked basically at any time as legitimate military targets, even if they constitute no direct threat to other parties to the conflict at the given point in time.

a) Principle of distinction: Is it possible to program autonomous systems in such a way that they are able to distinguish between protected civilians and enemy combatants, even in complex deployment scenarios?

Civilians are never legitimate targets, even in armed conflicts. Based on this premise is the central principle of distinction, laid down in particular in Article 51 of the first additional protocol to the Geneva Convention and also applicable under the common law of all states. Paragraph 2 of the provision makes it clear that neither the civilian population as such nor individual civilians may be subject of an attack. Several issues arise from this for the context of autonomous weapons systems. First, it has to be clarified whether the sensors of such systems could draw the necessary distinction with sufficient reliability. This is already a technological challenge that some robotics experts believe to be impracticable.<sup>66</sup>

Beyond this purely material capability of differentiation, however, account also has to be taken of the fact that complying with the imperative of distinction requires highly complex appraisal processes. In the critical situations of an armed conflict intricate value judgements

National Security and Law Essay Series, 2013, <http://media.hoover.org/sites/default/files/documents/Anderson-Waxman-LawAndEthics-r2-FINAL.pdf> (accessed on 11 February 2015).

66. See, for example, Noel E. Sharkey, *The Evitability of Autonomous Robot Warfare*, *International Review of the Red Cross*, Vol. 94, 2012, pp. 787, 788.

always arise. Even presupposing major advances in sensor technology the question remains whether this aspect could ever be handled by algorithms. That applies in particular to typical conflict situations in present-day armed conflicts, which are characterised by increasing confusion and complexity. Slogans such as »asymmetrical warfare« and »urban warfare« conceal the enormous difficulties involved in differentiating between irregular fighters and other legitimate targets, on one hand, and the civilian population – which must be protected – on the other. This is because this is not a matter of sensors' identifying particular weaponry or enemy uniforms, but the interpretation of human behaviour. It is unclear whether computer systems could ever be capable of complying with the imperative of distinction given this degree of confusion. It is noteworthy in this context that the guidelines of the US army concerning ethical conduct on missions expressly includes the criterion of »gut feeling«. Thus the soldier, before using his or her weapon, is supposed, as a last mental step, to decide whether a particular action is morally »right«. Even advocates of autonomous weapons systems acknowledge that such deliberation is not amenable to algorithms.<sup>67</sup> It appears particularly problematic to program rules of behaviour for equivocal combat situations. Thus some experts at the very least call into question that it is possible even in principle to equip autonomous weapons systems with a capability to »doubt« that could effectively prevent it from continuing with an attack in such a situation.<sup>68</sup>

On the other hand, human soldiers are already confronted with the same difficulty. They make mistakes and can consciously or unconsciously violate the principle of distinction. Stress, anger or fear are factors that can trigger or make more likely legal transgressions. It is precisely such emotions that machines do not have to cope with. This is raised as the main argument why autonomous weapons systems would be much more capable than human beings of complying with the principle of

distinction in difficult situations.<sup>69</sup> For a robot that does not have to fear for its life it is much easier to comply with the assumption demanded under international law that someone is a civilian, who must be protected, up to the point when they actually draw their weapon. A human soldier for the sake of self-preservation inherently has an overriding interest in reversing this assumption. At the same time, however, the question arises of why robots should be allowed to kill enemy combatants and operate under the same rules as human combatants in the first place. If robots are really so superior to human beings, as is often asserted, and given that (armoured) robots could fight without any significant risk to themselves, they should also be bound to correspondingly higher – in accordance with the current state of technology – safety standards. We shall return to and develop this question below.

Another consideration is the problem of so-called »scenario fulfilment«. This refers to the unconscious execution of a prerehearsed chain of actions based on a falsely perceived trigger factor, resulting in the lethal deployment of a weapon.<sup>70</sup> Such a loss of adequate perception of reality, which causes a soldier to blank out critical information, could not happen to an autonomous weapons system.<sup>71</sup>

It should also be noted that there is more at issue here than the protection of civilians. Combatants or enemy fighters who have surrendered or are wounded count as *hors de combat* and enjoy the protection of international humanitarian law. It is at least questionable whether autonomous weapons systems will be capable of recognising these distinction criteria reliably. For that to be possible they would have to be able to register gestures, facial expressions and emotions and to evaluate them correctly. That is a very tall order, to say the least.<sup>72</sup>

67. Ronald C. Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, Technical Report GIT-GVU-07-11, p. 51, <http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf> (accessed on 15 February 2015).

68. Cf. Michael N. Schmitt, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, *Harvard National Security Journal Features*, 2013, p. 16, <http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf> (accessed on 15 February 2015); for a contrary view, Arkin, see above note 67, p. 58, who starts from the assumption that one could program the principle »first, do no harm« into such a system, which would always prohibit it from automatically deploying deadly force if the situation cannot be classified unequivocally.

69. Arkin, see above note 67, p. 6.

70. The accidental shooting down of Iran Air 655 by the USS Vincennes in July 1988 is attributed to this phenomenon; see <http://en.wikipedia.org/wiki/Iran-Air-Flight-655> (accessed on 15 February 2015).

71. Arkin, see above note 67, p. 6.

72. Sassòli, see above note 54, p. 327.

b) Principle of proportionality: Is it possible to program autonomous systems in such a way that excessive »collateral damage« can be safely ruled out?

Besides the principle of distinction the principle of proportionality is another fundamental pillar of international humanitarian law. While the principle of distinction strictly prohibits direct attacks against civilians, indirect harm to the civilian population – for example, if a military objective is attacked directly and civilian bystanders are killed or injured – are permitted to the extent they are proportionate. This rule, which also applies for all states by virtue of customary international law, finds its textual expression in Article 51(5)(b) and Article 57(2)(a)(iii) of the First Additional Protocol to the Geneva Conventions. Specifically, it prohibits 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« (Article 51(5)(b)). The mental operation required for such deliberation between the anticipated military advantage and possible civilian damage involves complex, value-based case-by-case decision-making, in which the circumstances have to be weighed in their totality. Again, the question arises whether this can be performed by means of calculation using algorithms. Could an autonomous system in fact correctly assess the military advantages of an operation? The International Criminal Court for the former Yugoslavia laid down in 2003 that what is required to determine whether an attack is proportionate is to examine whether a »reasonably well-informed person«, under the circumstances prevailing at the time in question, and taking due account of the available information could have expected that the attack would result in an excessive number of civilian casualties.<sup>73</sup> Some experts call into question whether autonomous systems, at least in the near future, could be programmed to perform such an evaluation. Should this not prove to be the case or if autonomous systems could not distinguish with certainty between protected civilians and enemy combatants and fighters, such systems may from the very outset be deployed only in circumstances in which civilians are not present.<sup>74</sup>

73. ICTY, Prosecutor v. Stanislav Galic, Judgment (Trial Chamber) (Case No. IT-989-29-T), 5 December 2003, recital 58.

74. William H. Boothby, *Conflict Law: The Influence of New Weapons Technology*, Den Haag 2014, pp. 110 f.; Sharkey, see above note 66, p. 789.

Other commentators, in turn, counter this criticism with the argument that the requisite programming would presumably be possible to the extent that it would be equivalent to a human evaluation in the same situation.<sup>75</sup> Robotics scientist Ronald Arkin even goes one step further and points to autonomous systems' lack of a survival instinct. As a result, according to Arkin, they could not be misled into applying excessive force of arms, which means that compliance with the principle of proportionality could be ensured much more reliably.<sup>76</sup> However, this objection does not rebut the arguments of those who already doubt whether autonomous systems could at all be programmed in such a way that they could appropriately apply the principle of proportionality and comply with it accordingly.

c) Precautions: Is it possible to program autonomous systems in such a way that they are able to safely comply with humanitarian law obligations to take precautions in attack?

According to the third principle of international humanitarian law, autonomous weapons systems must ultimately also be capable of observing the humanitarian law principle of precautions. This is closely linked to the imperative of distinction and the principle of proportionality and, according to Article 57(1) of the First Additional Protocol, requires that »[i]n the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects«. Accordingly, measures must continually be taken to prevent civilian losses. The duty of care applies to the whole planning phase of an armed deployment and concerns all persons involved in preparations, in other words, commanders, but also arguably the manufacturers and programmers of systems.<sup>77</sup> But it goes even further. The original planning must also be valid and decisive after the mission has begun. Because many unforeseen things can happen in the course of a combat mission, some authors argue, the principle implicitly gives

75. Schmitt, see above note 68, p. 19.

76. Arkin, see above note 67, p. 58; Noel Sharkey criticises Arkin's approach generally, on the ground that it envisages only a back-end system, which relies on the development in the near future of sensors capable of providing such software with an adequate sense of the outside world. The development of such sensors is not on the horizon, however, which means that Arkin's argument is merely a thought experiment and is likely to remain so; Sharkey, see above note 66, p. 790.

77. Boothby, see above note 74, p. 115.

rise to the duty to keep a human soldier always at least »on the loop«, so that they can respond spontaneously to changed circumstances.<sup>78</sup> Others express doubts concerning this mode. Because computer-guided weapons systems process information much more rapidly than human beings and thus could react more quickly, it is questionable in how far one could assume that soldiers »on the loop« in combat situations are capable of intervening if the weapons system is preparing to violate a rule of international humanitarian law.<sup>79</sup>

Real »precaution« thus means deploying autonomous weapons systems only where an encounter with the civilian population is excluded from the outset. The actual decision with regard to the rules of international humanitarian law will then be taken by the commander who sends the machine into the field. The question remains, however, how realistic such a scenario is, given the forms of conflict prevailing today. Once autonomous weapons systems come into existence and become operational it will be difficult to avoid succumbing to the temptation to deploy them, even in complex and unpredictable situations. It is precisely in such situations that human soldiers are particularly under threat, making the incentive to replace them with robots particularly strong.

d) Preliminary findings on the question of whether autonomous systems can comply with the rules of international humanitarian law

All in all, much remains speculative with regard to whether autonomous weapons systems will be able to comply with the fundamental rules of international humanitarian law. Some experts expect algorithms and sensors to be able to provide the requisite capability, others remain doubtful. International humanitarian law imposes exacting requirements on the judgement and ability to interpret the most various critical situations in the field. Modern conflicts are characterised by a high level of unpredictability and complexity. Even the best autonomous systems will never possess an all-encompassing contextual intelligence. By their very nature, they have no possibility of thinking outside the box of their algorithms. That can prove to be fatal

if something unforeseen occurs. It is at the very least questionable whether robots can operate safely when they have to deviate from the original mission plan.<sup>80</sup>

On the other hand, if it actually turns out that autonomous systems are significantly more capable of complying with the imperative of distinction and the principle of proportionality than human soldiers, then it could even be argued, on the contrary, that the military commander is almost (at least morally) duty-bound to deploy autonomous systems instead of human beings to carry out a mission in conflict situations in which civilians are present. Because under such circumstances only in this way could it be ensured that civilians will be protected as far as possible. It is stressed repeatedly that today human beings are the weakest link in military decision-making procedures. Often, war crimes are the outcome of soldiers' emotional overload in the face of the complexity of events and fear on the battle field. In this respect autonomous weapons systems are always superior and thus in a better position to make ethically correct decisions. The problem of deliberation on the basis of moral-legal fundamental principles may therefore be exaggerated: it is not a matter of arriving at one's own judgement based on one's own deliberations. On the contrary, soldiers are only supposed to apply those rules that the international community has established on the basis of universally valid considerations. Precisely when the application of fixed rules is at issue, algorithms are typically more capable, for the reasons laid out above.<sup>81</sup>

A more fundamental consideration follows on from this, however, which has only been taken up in the current debate recently and goes far beyond what we have discussed so far: is not the whole system of international humanitarian law based on the implicit assumption that it concerns and is applied by human decision-makers in armed conflicts?<sup>82</sup> If the addressees of the established duties and prohibitions are human beings, whose emotionality, susceptibility to error and survival instinct are already factored in to the legal architecture, then the question of whether autonomous weapons systems can comply with international humanitarian law is erroneous.

78. Academy Briefing No. 8, see above note 11, p. 16.

79. Alston, see above note 62, p. 54.

80. Paul Scharre, *Why Unmanned*, *Joint Force Quarterly*, Vol. 61, 2011, pp. 89, 92; in Sharkey's words, »When a machine goes wrong it can go really wrong in a way that no human ever would«, see above note 66, p. 790.

81. Arkin, see above note 67, p. 55.

82. Asaro, see above note 64, p. 700.

More decisive may be the question of whether the rules contained in this legal order are still the right rules if autonomous combat systems are involved in warfare.

From this consideration one can conclude that such systems – if at all – should have to satisfy a much higher standard. Among other things, this might mean that, for example, the standard of protection contained in the principle of distinction – given the very simplistic distinction between directly targetable and protected persons – should be regarded as too weak. Instead, a legal duty could be established for the developers of autonomous weapons systems to program them in such a way that they use force only in the case of unequivocally aggressive and offensive behaviour on the part of enemy combatants/fighters. In situations, by contrast, that are not clear-cut in this respect such systems would have to refrain from the use of lethal force even if human soldiers in an identical situation would be permitted to reach for their weapons. Because autonomous systems can operate without existential risk to themselves it appears entirely reasonable – at least in scenarios in which such systems encounter human beings directly – to require a significantly longer delay and to hold them to a higher legal standard of protection, compliance with which could not be expected from a human being in what for them would be a potentially life-threatening situation. It should always be determined, therefore, whether under the given circumstances it is absolutely necessary that a human being loses his or her life.<sup>83</sup> This consideration could even be taken so far as to require, as a last resort, that autonomous weapons systems may not apply deadly force, but rather only incapacitate an enemy by non-lethal measures. These general considerations, which go beyond the existing regulatory framework of international humanitarian law, will be analysed in more detail below.

### 3. Do Lethal Autonomous Systems Violate Human Dignity?

Even more fundamental is the question of whether it may be a fundamental – that is, regardless of whether autonomous systems could comply with the currently applicable rules of international humanitarian law – violation of human dignity to entrust the decision on

whether to kill to a machine. In this context it is, first of all, important to recognise that the protection of human dignity has a different status in international law from the one it has in Germany's Basic Law. Although the principle is acknowledged as an ethical maxim of the international law regime, legal status and substance, by contrast, are far less clear-cut than under German legislation.<sup>84</sup> In any case, human dignity is not necessarily regarded as absolute and *per se* »inviolable« everywhere in the world and in the same way as in Article 1 of Germany's Basic Law. Sometimes it is also seen as an important consideration that is to be balanced against other considerations. In Article 8(2)(b)(xxi) and 8(2)(c)(ii) of the International Criminal Court's Rome Statute it is, after all, stipulated that an action that violates personal dignity meets the conditions categorising a war crime. And even if the standard and definition of German constitutional law cannot be applied, human dignity, in philosophical terms, means that each person must be regarded as an individual and treated accordingly, as a unique and irreplaceable being. It follows from that, regardless of purely legal provisions for the deployment of autonomous systems in armed conflicts, that there is an ethical dimension that has to be taken into consideration.

Is this principle of the unique worth of human life violated when decision-making about life and death is ceded to a totally »rational« – that is, operating on the basis of algorithms – machine? It appears to be at least doubtful whether this procedure is, ultimately, entirely rationalisable. The inherent irrationality that is always part and parcel of a human decision to kill could itself be regarded as a prerequisite for at least a minimum degree of moral substance. Because even if a soldier has the right, in accordance with the principles of international humanitarian law, to kill an enemy combatant in a specific situation, such an action, even when the appropriate state of command is in place, is always preceded by a highly personal examination and decision arising from one's conscience.<sup>85</sup> Such an operation of human reason, which includes the power of judgement and empathy as causes or motives of action, is not available

84. Niels Petersen, *Human Dignity, International Protection*, in Rüdiger Wolfrum (ed.), *Max Planck Encyclopedia of Public International Law*, Oxford 2012, marginal number 1.

85. O'Connell, see above note 6, p. 231; on this cf. the – controversial – studies of US soldiers in the Second World War, a comparatively small proportion of whom supposedly ever fired a weapon; see Dave Grossman, *On Killing: The Psychological Cost of Learning to Kill in War and Society*, Boston 1995.

83. *Ibid.*, p. 701.

to autonomous weapons systems. The functioning of such systems, by contrast, makes it possible to make decisions on killing literally with merciless consistency, without prior moral deliberation.<sup>86</sup> A human being is not then regarded as an individual but as a mere object of a mathematically calculated decision to kill. UN Special Rapporteur Heyns in this context aptly talks of »death by algorithm«. <sup>87</sup> Against this background it appears entirely justifiable to classify machine decision-making on life and death – not autonomy in military systems per se – as a violation of human dignity. It should be underlined in this context, however, that the German Constitutional Court – by international comparison certainly one of the staunchest defenders of human dignity – in its judgment on the Aviation Security Act implicitly set the threshold for a violation of human dignity in the particular situation of armed conflict comparatively high.<sup>88</sup>

Furthermore, a person attacked by an autonomous weapons system basically lacks the opportunity to appeal to the attacker's humanity. Factors such as dignity or empathy are removed from the equation. Robots have no understanding of the inherent value of human life. Killing is not always necessary. Even if such systems could be programmed in such a way that they always apply deadly force within what is legally permissible the possibility exists that people will die who otherwise would have been spared.

If in future combat robots were to be deployed in areas with a resident civilian population it should also be noted that this could impair people's capability of living a dignified life. UN Special Rapporteur Heyns to that extent was right to point out that the deployment of unmanned, autonomously operating weapons systems amidst civilians could cause general trepidation, anxieties or traumas among them.<sup>89</sup> The major consequences of

the continuing deployment of drones by the United States in this regard have been investigated in detail and the findings are disturbing.<sup>90</sup> Under such circumstances it is scarcely possible to live a normal everyday life. It is plausible, to say the least, that the presence of potentially lethal combat machines could have a similar effect.

#### 4. Is There an Obligation to Design Autonomous Weapons Systems to Act in a Non-lethal Manner?

In view of the preceding arguments it follows that there could be a duty to build or deploy autonomous weapons systems only in such a way that they are unable to kill human beings – whether civilians or combatants. As already remarked, the absolutely fundamental question arises of whether the principles and value judgements underlying the current law on armed conflict can still find application to this completely new kind of weapon.

The conflict party that deploys robots acts without risk to its own soldiers. If one starts from the premise that killings in war are justified (solely) by the reciprocity of the killing, then this justification of lethal actions is eliminated. The extent to which this argument applies is debatable, however. After all, armed conflicts have been characterised for years by the very asymmetrical initial situations in which, due to technological superiority, there is often no acute, direct personal risk to one particular side. This applies not only to drone deployments in the so-called »war on terror«. Even during the air attacks – often conducted from great height and beyond the reach of the enemy – launched by NATO against Serbia in 1999 the allied pilots were not exposed to any significant, immediate danger. It would be inappropriate to talk of reciprocity in this instance. In the history of weapons technology debates on the implications of new kinds of weapons or new methods of warfare made possible by technological developments have always revolved around the question of whether the risk minimisation arising from them is ethically defensible or not.<sup>91</sup> International humanitarian law in any case does not expressly prohibit reducing the risk to one's own soldiers by means of

86. Asaro, see above 64, p. 695.

87. Comments by Christof Heyns, UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Informal Meeting of Experts on Lethal Autonomous Weapons: Conventional Weapons Convention, 16 April 2015, p. 5, Comments by Christof Heyns, UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Informal Meeting of Experts on Lethal Autonomous Weapons: Conventional Weapons Convention, 16 April 2015, p. 5, [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/1869331AFF45728BC1257E2D0050EFE0/\\$file/2015-LAWS-MX-Heyns-Transcript.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/1869331AFF45728BC1257E2D0050EFE0/$file/2015-LAWS-MX-Heyns-Transcript.pdf) (accessed on 28 April 2015).

88. For criticism of this, see Andreas Zimmermann/Robin Geiß, Die Tötung unbeteiligter Zivilisten: Menschenunwürdig im Frieden – menschenwürdig im Krieg? [The killing of non-participating civilians: inhumane in peace, humane in war?], in *Der Staat* 46 (2007), pp. 377–393.

89. Heyns, see above note 5, marginal number 98.

90. International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU School of Law, Living under Drones: Death, Injury, and Trauma to Civilians From US Drone Practices in Pakistan, 2012, <http://www.livingunderdrones.org/wp-content/uploads/2013/10/Stanford-NYU-Living-Under-Drones.pdf> (accessed on 15 February 2015).

91. Anderson/Waxman, see above note 65, p. 8.

weapons technology.<sup>92</sup> To the contrary, fairness is not a relevant category of international humanitarian law. In fact, it is hardly possible to construe an ethical obligation to put the lives of one's own military personnel in danger. To that extent, the argument, taken in isolation, ultimately is scarcely convincing.<sup>93</sup>

On the other hand, it might be considered that at least in the event of the exclusive deployment of unmanned autonomous systems it no longer makes sense to talk of »war«.<sup>94</sup> If one pursues this line of argument, in addition to the regulatory standards of international humanitarian law, (stricter) human rights standards could be adduced to regulate autonomous weapons systems. This idea sometimes encounters significant criticisms in the international community, however. In particular the United States and the United Kingdom recently, on the occasion of an informal expert meeting under the aegis of the UN Convention on the prohibition or restriction on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects (UN Weapons Convention), in April 2015, once more underlined that autonomous weapons systems should be evaluated solely in accordance with the rules of international humanitarian law.<sup>95</sup> A corresponding discussion is already under way on drones and the »war on terror«. If the right to life guaranteed by human rights law was to apply without restriction to the deployment of combat robots – in other words, without prejudice to the special measures of the law of armed conflict – then the killing of human beings would be permitted only as an absolute exception. The ninth of the UN Basic Principles on the Use of Force and Firearms by Law Enforcement Officials of 1990 clarifies the narrow limits.<sup>96</sup> It states that armed force may be used against persons only if there is an imminent threat

of death or serious injury either to the law enforcement official him- or herself or to a third person, or to prevent the perpetration of a particularly serious crime involving grave threat to life. Even then, a weapon may be drawn only if no less drastic option is available. Deadly force may be used intentionally only if there is absolutely no alternative to protect human life. Because a robot does not have a life of its own (worthy of protection) then only the danger to the life of a third person could come into consideration.<sup>97</sup> Apart from these very restricted exceptional situations the application of force would be proportionate and justifiable only if it does not lead to the killing of the enemy.

But even if one regards only the regulatory system of international humanitarian law to be applicable it is – at least potentially – possible to derive an obligation to deploy combat robots only as non-lethal systems. The International Committee of the Red Cross, in a 2009 study, derived the basic principle that in armed conflict the deployment of lethal force against non-protected persons – for example, enemy fighters – is inherently limited by what is necessary in the given circumstances. In other words, even in armed conflict one should not proceed with more violence than is absolutely necessary in the given situation in order to achieve a legitimate military goal.<sup>98</sup> Accordingly, lethal force should be used against the military opponent only if no less drastic means (capture) are available. The study argues that the mere fact that a person is not protected against the use of force under international humanitarian law does not automatically imply that they may be killed.<sup>99</sup> As clear and plausible as this proposal may sound, it was revolutionary given the accepted understanding of international humanitarian law. It remains highly controversial and to date has not been accepted – as far as can be seen – as legally binding by states.<sup>100</sup> Nevertheless, this approach is

92. Schmitt/Thurnher, see above note 22, p. 281.

93. Cf. also Münkler, see above note 20.

94. For example, UN Special Rapporteur Heyns, see above note 5, marginal number 60.

95. The US opinion of 13 April 2015 is available at: [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/8B33A1CDBE80EC60C1257E2800275E56/\\$file/2015-LAWS-MX-USA+bis.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/8B33A1CDBE80EC60C1257E2800275E56/$file/2015-LAWS-MX-USA+bis.pdf) (accessed on 27 April 2015). It says that: »We believe our discussion here in CCW, a forum focused on international humanitarian law, remains the relevant framework for this discussion.« The opinion of the United Kingdom of 13 April can be found at: [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/1CBF996AF7AD10E2C1257E260060318A/\\$file/2015-LAWS-MX-United+Kingdom.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/1CBF996AF7AD10E2C1257E260060318A/$file/2015-LAWS-MX-United+Kingdom.pdf) (accessed on 27 April 2015).

96. The Basic Principles are available at: <http://www.ohchr.org/EN/ProfessionalInterest/Pages/UseOfForceAndFirearms.aspx> (accessed on 15 February 2015).

97. Sassòli even derives from the Basic Principle that robots in general may be deployed only if there is a possibility of making an arrest; see above note 53, p. 318. It is of course questionable whether it could always be guaranteed before a mission that the deployment of armed force will not be necessary.

98. Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, Geneva 2009, p. 77 ff., <https://www.icrc.org/eng/assets/files/other/icrc-002-0990.pdf> (accessed on 15 February 2015).

99. *Ibid.*, p. 78 f.

100. Cf., for example, W. Hays Parks, Part IX of the ICRC *Direct Participation in Hostilities Study: No Mandate, No Expertise, and Legally Incorrect*, *New York University Journal of International Law and Politics*, Vol. 42, 2010, p. 769; on the author's reply, see Nils Melzer, *Keeping the Balance between Military Necessity and Humanity: A Response to Four*

finding increasing support and dissemination at least in the literature and occasionally even in practice.<sup>101</sup>

Based on this approach one might argue that for an (armoured) robot, which in combat with human opponents is undoubtedly far superior, there pretty much never is an urgent necessity – at least in scenarios involving a direct confrontation between human and machine that entail the possibility of capture – to apply lethal force, particularly because a robot has no life of its own to put into danger. That being the case, an autonomous weapons system in such situations should always attempt merely to incapacitate or capture the enemy. It would appear entirely reasonable that in future robots should be deployed in high-risk operations, such as house searches in conflict areas. In such scenarios the reduction of potentially lethal force to situations of urgent necessity would appear to be imperative.

## 5. The Problem of Liability

Following on from substantive ethical and legal issues is the problem of liability. If an autonomous weapons system violates international law and possibly even meets the criteria for designating an action a war crime then on whom and on what basis does liability lie? Conceptually, it makes no sense to lay it on the systems themselves. Even if we proceed from the assumption that they are genuinely »intelligent«, any notion of liability that notionally implies some kind of sanction is misconceived from the very outset.<sup>102</sup>

Liability is fundamental as a legal basis for protection guarantees, both in international law and with regard to human rights.<sup>103</sup> In what follows we shall distinguish between individual criminal law, civil law and state liability. The respective areas give rise to a plethora of complex legal issues, which we can outline here only briefly.

Critiques of the ICRC's Interpretive Guidance on the Notion of Direct Participation in Hostilities, *New York University Journal of International Law and Politics*, Vol. 42, 2010, p. 831.

101. R. Goodman, The Power to Kill or Capture Enemy Combatants, *European Journal of International Law*, Vol. 24, No. 3 (2013), pp. 819–853.

102. Heyns, see above note 5, marginal number 76.

103. Human Rights Watch, see above note 27, p. 42.

### a) Criminal Responsibility

With regard to the question of who could be held to account under criminal law if the actions of an autonomous weapons system violate international law two starting points suggest themselves: (i) the programmers or manufacturers of the system, on one hand, and the commanders in the field, on the other. In addition, at least in certain cases superior officers, such as generals, or even political decision-makers who took the fundamental decision on deployment can come into consideration when it comes to attributing liability.<sup>104</sup> First of all, the legal position is fairly clear, insofar as a deliberate war crime is in question. If, then, a programmer deliberately writes algorithms in such a way that the robot subsequently attacks civilians in the field, then criminal responsibility can easily be established. A similar conclusion applies in case of a commander who is aware of the machine's shortcomings, but nevertheless deploys it in an area with a high proportion of resident civilians because the civilian casualties that will inevitably occur are of no concern to him. Difficulties start to arise only when all the human beings involved assume that the weapons system functions properly and do not deliberately set out to kill protected persons.

Autonomous weapons systems are extremely complex. It is not always easy even for those involved in their manufacture to estimate from the outset all possible consequences of their deployment. The point of such systems is that they should be able to react autonomously to unforeseen situations. If one takes the notion of autonomous action seriously, the testing of all possible modes of behaviour that might become an issue during a combat mission is excluded in principle. In the field, environmental conditions and influences that could not have been foreseen continually come into play.<sup>105</sup> In criminal proceedings, however, such factors inevitably have an impact on the programmer's individual liability because predictability is a prerequisite of criminality even in the case of offences of negligence. Then, however, there is a danger that ultimately all misconduct on the part of autonomous systems presents itself as force majeure, in other words, as an event that could not have

104. Heyns, see above note 5, marginal number 77.

105. US Chief Air Force Scientist, Report on Technology Horizons: A Vision for Air Force Science and Technology during 2010–2030, 2010, p. 105 f., <http://www.flightglobal.com/assets/getasset.aspx?itemID=35525> (accessed on 16 February 2015).

been prevented. This problem is exacerbated if adaptive algorithms are involved. In that case it is even more difficult to anticipate in advance how an autonomous weapons system will behave.

The criminal responsibility of the commander is also problematic. One starting point would be an appropriate application of Article 28 of the Rome Statute of the International Criminal Court, which regulates the international criminal law liability of military commanders. This requires, however, that the commander at the given point in time knew or should have known that the subordinate was committing or about to commit a crime. Basically, two problems arise here. First, this provision cannot be applied analogously without further ado. It is based on the premise of a relationship between autonomous and independently acting persons. This construction cannot be transferred to the relationship between human and robot. This is already demonstrated by the abovementioned technological considerations: if algorithms make the behaviour of a weapons system unpredictable or, moreover, are even adaptive, at what point can it be meaningfully said that the officer »should have known« that the autonomous system would violate rules of international humanitarian law?<sup>106</sup> One can arrive at another assessment, however, if the commander observes as the robot is beginning, due to software errors or other malfunctions, to commit a war crime. If he does not promptly abort the mission by deactivating the machine he is at least criminally liable for all transgressions from that point on.<sup>107</sup>

This point shows that liability decreases where the autonomy of the system increases. This is a structural problem: liability is conditional on control. The more autonomously a system is able to act, the greater the potential liability gaps. Criminality can no longer be attributed from a certain point onwards. That means that the problem cannot be circumvented simply by reference to necessary human involvement at a certain point in time.<sup>108</sup> The auxiliary argument that individual criminal responsibility is in any case exaggerated because this is only one way among several of ensuring compliance with

the rules of international humanitarian law<sup>109</sup> mistakes the fundamental meaning of the legal institution of criminal liability in this respect. Insofar as autonomous weapons systems come to be deployed in the future existing gaps in the attribution of criminal responsibility should be closed focussing on the tremendous dangers arising from the deployment of autonomous (offensive) weapons systems. On this basis, for example, if the dangerousness of these systems found expression in a violation of international law the criminal liability of the commander, who created the hazardous situation by deploying the system, could be derived automatically.

#### b) Civil Liability

Civil liability must be distinguished from individual criminal responsibility. In this connection some authors advocate strict liability in the sense of strict product liability, which affects the manufacturers of autonomous weapons systems.<sup>110</sup> Such a high standard of liability could cause manufacturers to ensure that the safest possible systems are produced.<sup>111</sup> However, UN Special Rapporteur Christof Heyns points out that such rules on robot technology have not yet been established on the domestic level. Furthermore, civil law proceedings have the major disadvantage that they impose the obligation on victims to make a complaint. Given the nature of conflict situations today potential victims will rarely be in a position to initiate proceedings against the manufacturers of autonomous weapons systems, who presumably are located in industrialised countries.<sup>112</sup> In 2013 the German Constitutional Court, in its ruling on the NATO attacks on the Varvarin bridge in Serbia, found that the existence of »a general international customary law regulation in accordance with which individuals, in the event of a violation of international humanitarian law, have a claim to damages or compensation against the responsible state ... cannot be established, at least at present«.<sup>113</sup>

106. Asaro, see above note 64, p. 693; Sharkey, see above note 65, p. 790.

107. Needless to say, that requires that the commander is at least in some respect still »on the loop«, in other words, has a real opportunity to intervene and abort the mission.

108. But see Schmitt/Thurnher, see above note 22, p. 277.

109. Anderson/Waxman, see above note 65, p. 17.

110. Heyns, see above note 5, marginal number 79.

111. Human Rights Watch, see above note 27, p. 44.

112. Heyns, see above note 5, marginal number 79; Human Rights Watch, see above note 27, p. 44.

113. BVerfG, 2 BvR 2660/06, 2 BvR 487/07, of 13 August 2013.

### c) State Responsibility

What remains to be clarified is the question of state responsibility in cases in which a country's armed forces deploy autonomous weapons systems. According to Article 4 of the ILC Articles on State Responsibility states are responsible for the actions of their organs.<sup>114</sup> An extension or appropriate application of this provision also to autonomous systems would seem to go too far. Members of a state's regular armed forces certainly come within the purview of this norm, however. The deployment of an autonomous weapons system by a military commander is undoubtedly to be considered an action in this sense – thus the responsibility of a state cannot be brushed aside merely by referring to the wording of Article 91 of the first additional protocol to the Geneva Convention, which expressly only talks about a conflict party being responsible for actions »that are committed by *persons* belonging to their armed forces«. <sup>115</sup> If an autonomous weapons system commits a war crime on a mission or otherwise violates international law then that is to be attributed – at least in the first instance – to the state whose armed forces deployed the device.

Another problem is the question of whether state responsibility is conditional on whether the direct action was committed in such a way that individual liability could be adduced, in other words, that it was intentional or negligent. If yes, then in the case of state responsibility one would find oneself confronted with the same difficulties as in the case of individual criminal responsibility. Whether the distinguishing feature of individual responsibility comes within the scope of the rule is not determined on the secondary level of state responsibility itself. Rather it depends on whether this is provided for by the primary norm.<sup>116</sup> Although in the normal case responsibility for violations of the regulations of international humanitarian law is linked to

the condition that the acting combatants can be blamed for their deeds, the commentary of the International Committee of the Red Cross on the first additional protocol of the Geneva Conventions makes it crystal clear that, subject to a corresponding regulation, there can also be strict liability regardless of fault if injury occurs as a result of a dangerous action attributable to the state.<sup>117</sup>

Such strict liability appears to be a feasible approach at least with regard to the responsibility of states for the actions of autonomous weapons systems. The deployment of robots whose decision-making and modus operandi are not subject to the direct and complete control of human beings is inherently risky. Liability for such a calculated risk in relation to actions that in themselves are not prohibited (in relation to autonomous weapons systems this would be the decision to deploy the systems) is gradually being established in more and more areas in international law. For example, in Article II of the 1972 UN Convention on International Liability for Damage Caused by Space Objects:<sup>118</sup>

*A launching state shall be absolutely liable to pay compensation for damage caused by its space object on the surface of the earth or to aircraft in flight.*

Similar to space objects in the first decades after the Second World War – thus in the first instance satellites in orbit around the earth – autonomous (combat) robots will in future constitute a technological leap whose consequences we can still only dimly imagine. Strict liability would appear to be in order in this case. Such a liability model would require a new international agreement, however, given the current state of international law. Despite increasing state practice responsibility for in itself legitimate, but risky actions giving rise to damage is still not – and particularly not in the context of armed conflict – considered to be a norm under international customary law.<sup>119</sup>

If an autonomous weapons system causes damage in the course of a mission, besides strict liability a genuine breach of duty on the part of the deploying

114. <http://legal.un.org/ilc/texts/instruments/english/draft%20articles/9-6-2001.pdf> (accessed on 16 February 2015).

115. Emphasis added; but see John Frank Weaver, *Asimov's Three Laws Are Not an International Treaty: How to Make Treaties Govern »Killer Robots«*, *Slate*, 1 December 2014, <http://www.slate.com/articles/technology/future-tense/2014/12/autonomous-weapons-and-international-law-we-need-these-three-treaties-to.html> (accessed on 16 February 2015).

116. International Law Commission, *Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries*, Article 2, marginal number 3, <http://legal.un.org/ilc/texts/instruments/english/commentaries/9-6-2001.pdf> (accessed on 16 February 2015).

117. International Committee of the Red Cross, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*, Geneva 1987, Article 91, marginal number 3661, <http://bit.ly/1CEeY1p> (accessed on 16 February 2015).

118. <http://www.admin.ch/opc/de/classified-compilation/19720066/20111210000/0.790.2.pdf> (accessed on 16 February 2015).

119. Knut Ipsen, *Völkerrecht*, 6th edition, Munich 2014, p. 559 f.

state also comes into play. This is to be assumed if the state has not complied with its – to be defined in more detail – duties with regard to prevention, monitoring and damage prevention in relation to autonomous weapons systems. In accordance with Article 1 of the four Geneva Conventions of 1949 the states parties are obliged to comply with the Geneva Conventions and to ensure compliance under all circumstances («to respect and to ensure respect for the present Convention in all circumstances»). The due diligence obligation to ensure respect arising from this provision in principle appears transferable to the deployment of autonomous weapons systems.<sup>120</sup> In this context due diligence could mean – because this is an issue for the future and there is no state practice to speak of, specific determinations are not yet possible – in particular legal regulation of the manufacturing process and testing the weapons before they are made operational as intensively as possible under realistic deployment conditions. A duty to continuously monitor and control the actual use of systems in the course of armed conflicts can also be derived from this overarching obligation to ensure respect.

## VI. Proposals for Regulating Autonomous Weapons Systems

Following on from the analysis of the legal and ethical implications of the deployment of autonomous weapons systems we shall briefly present the most important contemporary proposals for their containment.

### 1. Banning Autonomous Weapons Systems

Some NGOs are calling for autonomous weapons systems to be banned under an international agreement. They include, in particular, Human Rights Watch<sup>121</sup> and the »Campaign to Stop Robot Killers«<sup>122</sup> initiative. A few states – namely Bolivia, Cuba, Ecuador, Egypt, Ghana and Pakistan, as well as the Vatican and Palestine – support a ban. Some experts argue that there are ethical and legal duties to prevent autonomous weapons systems from ever being given the capability of deciding on

human life and death.<sup>123</sup> The justification for this is based on the points made above: it cannot be ensured that autonomous weapons systems would be in a position to comply with the regulations of international humanitarian law. The further dehumanisation of war would further lower the threshold for states with regard to commencing armed conflicts. The problem of the responsibility gap cannot be solved; the looming liability loopholes can be countered only by means of a total ban. According to Human Rights Watch these problems arise not just with regard to completely autonomous systems, but also in relation to weapons that have so much influence over the selection of targets that human controllers in practice are no longer able to exercise real control and thus are effectively »out of the loop«.<sup>124</sup>

Critics of a complete ban maintain that, particularly at this early stage of development, it would be irresponsible to ban autonomous weapons systems. They offer an opportunity to drastically reduce the risk faced by soldiers on one's own side, which is certainly valid consideration under international humanitarian law.<sup>125</sup> Furthermore, a ban also makes no sense because the development and deployment of such systems is already inevitable since the necessary technology basically already exists.<sup>126</sup> The advocates of a ban vehemently reject this argument.<sup>127</sup> Technological progress is by no means inevitable and there have already been new kinds of weapons – such as blinding laser weapons<sup>128</sup> – that were banned before they were deployed. On the other hand, as far as can be seen this is the only example of a weapons system on which such far-reaching agreement has been reached.<sup>129</sup> Outside the context of new weapons, however, reference is made to the ban on human cloning that still prevails in most countries, which is widely regarded as incompatible with human dignity.<sup>130</sup>

120. See, for example, Robert P. Barnidge, Jr., *The Due Diligence Principle under International Law*, *International Community Law Review*, Vol. 8, 2006, p. 81.

121. See above note 27.

122. See above note 14.

123. Asaro, see above note 64, p. 689.

124. Human Rights Watch, see above note 27, p. 46.

125. Schmitt/Thurnher, see above note 22, p. 281.

126. Anderson/Waxman, see above note 65, p. 27.

127. One only has to look at Asaro, see above note 64, p. 404 f.

128. O'Connell, see above note 6, p. 233; they were banned in 1995 by CCW Protocol IV, <https://www.icrc.org/applic/ihl/ihl.nsf/0/49de65e1b0a201a7c125641f002d57af?OpenDocument> (accessed on 17 February 2015).

129. Geiß, see above note 44, p. 240.

130. O'Connell, see above note 6, p. 233; cf., for example, the United Nations Declaration on Human Cloning, A/Res/59/280, 23 March 2005, <http://www.nrlc.org/uploads/international/UN-GA-DeclarationHumanCloning.pdf> (accessed on 17 February 2015).

## 2. Moratorium

In his report to the UN Human Rights Council UN Special Rapporteur Christof Heyns takes the view that the issue requires an international negotiation process in order to find answers to the questions arising from autonomous weapons systems. For this purpose an interdisciplinary committee should be set up at the level of the United Nations. The aim should be a consensual international legal framework for the development and deployment of autonomous weapons systems.<sup>131</sup> Until such a framework is set up there should be moratoria at national level on the testing, manufacture, sale and deployment of such systems.<sup>132</sup> Bioethicist Wendell Wallach is also calling for such a moratorium.<sup>133</sup> As far as can be seen, however, no state has yet implemented such a moratorium.

## 3. Requirement of »Meaningful Human Control«

There is currently reason to believe that the requirement of »meaningful human control« could establish itself as a substantively open, but constructive starting point for further debate. There appears to be a consensus among various states and many experts that autonomous weapons systems should be permitted only to the extent that meaningful human control is retained in relation to critical decisions, above all the decision to deploy deadly force. That at least has been the overwhelming thrust of the two informal expert meetings held so far, convened by the states parties of the UN Weapons Convention in May 2014 (under a French chair) and in April 2015 (under a German chair).<sup>134</sup> In their closing declarations international and non-governmental organisations backed this demand, particularly noteworthy being the »Campaign to Stop Killer Robots«,<sup>135</sup> the International Committee of the Red Cross<sup>136</sup> and the International Committee for Robot Arms Control.<sup>137</sup> On the occasion of the meeting in April 2015 Germany reaffirmed its position, according to which the decision to deploy

armed force – in particular decisions on life and death – must not be taken exclusively by an autonomous system without any possibility for human intervention with regard to selection of target and engagement.<sup>138</sup> The United Kingdom, France and Japan also announced that, for the time being, they have no wish to develop completely autonomous weapons systems. Instead, they merely want to enhance automatisation in order to improve the effectiveness of systems when deployed.<sup>139</sup>

What is meant exactly by »meaningful human control« remains (completely) open. This very openness harbours an opportunity, however. It would appear that numerous states are getting used to the idea of continuing the current discussion in light of this criterion, which, depending on how it is implemented, could contain more or less far-reaching prohibitions. Ultimately, the criterion means nothing other than a ban on complete autonomy in certain areas. Because where there is human control, by definition there cannot be complete autonomy. The criterion of »meaningful human control« is thus to a certain extent contradictory. But it does serve as a useful hook and constructive starting point for further substantive discussions on which critical decisions should always be subject to such control and how such »meaningful human control« should be implemented in detail. Thus the requisite level of control can refer to several factors: the time-span between the last decision taken by humans and the exertion of force by the machine; the environment in which the machine comes to be deployed, especially with regard to the question of whether civilians are present in that environment; the orientation of the deployment, in other words, whether the machine is supposed to engage in defensive or offensive tasks; the question of whether the machine is set up to apply lethal force; the level of training of the persons tasked with exercising control over the machine; the question of the extent to which people

131. Heyns, see above note 5, marginal number 114.

132. Ibid., marginal number 113; this demand has also been taken up by, for example, Austria; see Academy Briefing No. 8, see above note 11, p. 7.

133. Wallach, see above note 9.

134. <http://bit.ly/1vT2Xpr> (accessed on 16 February 2015).

135. <http://bit.ly/19zazTG> (accessed on 17 February 2015).

136. <http://bit.ly/1EqpHfM> (accessed on 17 February 2015).

137. <http://bit.ly/1A7FeBm> (accessed on 17 February 2015).

138. General Statement by Germany, CCW Expert Meeting Lethal Autonomous Weapons Systems, Geneva, 13–17 April 2015, p. 2, [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/97636DEC6F1CBF56C1257E26005FE337/\\$file/2015-LAWS-MX-Germany.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/97636DEC6F1CBF56C1257E26005FE337/$file/2015-LAWS-MX-Germany.pdf) (accessed on 27 April 2015): »We will not accept that the decision to use force, in particular the decision over life and death, is taken solely by an autonomous system without any possibility for a human intervention in the selection and engagement of targets.«

139. UK Ministry of Defence, Development, Concepts and Doctrine Centre, The UK Approach to Unmanned Aircraft Systems, Joint Doctrine Note 2/11, 30 March 2011, marginal number 508, [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/33711/20110505JDN-211-UAS-v2U.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505JDN-211-UAS-v2U.pdf) (accessed on 17 February 2015).

are in a position to intervene, should the need arise, and to halt the mission; the implementation of safeguards with regard to responsibility, for example, by complete electronic recording of the robot's actions.

#### 4. Curbing by Programming: The »Ethical Governor«

Some robotics scientists, above all Ronald Arkin, assume, by contrast, that a ban on autonomous weapons systems would not be wise because in the medium term the technology will develop in such a way that it will be much easier for them to comply with the rules of the law of armed conflicts than for human beings. For that reason it would be ethically irresponsible to ban these systems prematurely.<sup>140</sup> In order to achieve this standard Arkin has come up with the model of the »ethical governor«, namely, an algorithm that embeds a feedback loop into the control software of the autonomous weapons system, which either allows or forbids the system to deploy its weapons in particular instances.<sup>141</sup> A »review authority« would thus be programmed into the system that would evaluate legal appropriacy before any exercise of lethal force on the basis of the rules of international humanitarian law. Although Arkin does not rule out that, even with such an »ethical governor«, robots will make mistakes, he does assume that the error rate would be significantly lower than that of human soldiers.<sup>142</sup>

#### 5. Approach Taken by the US Defense Department

As early as 2009 the US Defense Department presented a first strategy paper that addresses fundamental procedures for autonomous weapons systems.<sup>143</sup> The approach is primarily pragmatic. Right at the start the directive postulates that autonomous and semi-autonomous weapons systems are to be conceived in such a way that there should always be »appropriate levels of human judgment over the use of force«.<sup>144</sup>

140. Anderson/Waxman, see above note 65, p. 21.

141. Arkin, see above note 67.

142. Other scholars have tried to show that Arkin's notion suffers from conceptual errors that ultimately cannot be eliminated; see, for example, Andreas Matthias, *Is the Concept of an Ethical Governor Philosophically Sound?*, 2011, <https://www.academia.edu/473656/Is-the-Concept-of-an-Ethical-Governor-Philosophically-Sound> (accessed on 17 February 2015).

143. U.S. Department of Defense, see above note 7.

144. *Ibid.*, p. 2.

Even though substantively this standard is not entirely unambiguous, it can at least be assumed that it is weaker than the demand for »meaningful human control«. UN Special Rapporteur Christof Heyns has rightly pointed out that »appropriate« in certain situations might mean exercising no control whatsoever over the deployment of a weapon.

Overall, the US Defense Department focuses less on specific restrictions of possible kinds of deployment than on imposing far-reaching and strict requirements on the inspection of systems' hardware and software. The directive sets out the conduct of intensive series of tests. Operators of autonomous and semi-autonomous systems are obliged to comply with the rules of armed conflict and all other applicable international law treaties. It is notable, however, that it is expressly laid down that only semi-autonomous systems may be deployed to exercise lethal force. Autonomous systems may not be used to select human targets even if they are supervised by a human being.<sup>145</sup> However, the document does not fundamentally rule out the development of weapons that could be deployed beyond these limitations, if they satisfy stricter technical guidelines.

Wendell Wallach rightly points out, in relation to the provisions on series of tests for autonomous systems, that military necessity could easily mean that such weapons are deployed before the envisaged tests have been completed. This happened, for example, in the wars in Iraq and Afghanistan, in which new technologies were deployed – such as unmanned drones – before all the tests that had been laid down had been carried out. Something similar could happen with autonomous weapons systems.<sup>146</sup>

#### 6. Making Existing Regulations More Specific with Regard to Autonomous Weapons Systems

The inspection mechanisms under international humanitarian law for the introduction of new weapons technologies are, in themselves, applicable to autonomous weapons systems. From Article 36 Additional Protocol I and the Martens clause it follows that new weapons may not infringe the basic principles of international

145. *Ibid.*, p. 3.

146. Wallach, see above note 9.

humanitarian law and other applicable international law rules. Against this background, various states, on the occasion of the Geneva expert meeting in April 2015, advocated further development and specification of the substantive requirements concerning the testing of newly developed weapons within the meaning of Article 36 Additional Protocol I, especially with regard to autonomous weapons systems.<sup>147</sup> More detailed substantive elaboration of the inspection mechanism of Article 36 Additional Protocol I would undoubtedly be helpful and desirable. However, because such inspections would primarily concern the question whether new weapons systems are compatible with the relevant international humanitarian law it is also clear that such inspections alone cannot deal with all of the various the ethical and international law issues and problems thrown up by autonomous weapons systems.

## 7. Laying Down New Rules and Higher Standards of Protection

If one follows the premise that the existing legal regulations and regulatory systems were conceived (only) for human actors, with their particular weaknesses and deficiencies, the question must arise of the extent to which combat robots that are able to act independently can fall under this regime. The norms regarding the conduct of hostilities are based on the basic ethical assumptions of over 100 years ago. Even if it can be argued that, to date, they have proved more or less up to the job as new kinds of weapons have been developed, when it comes to combat artificial intelligence it would appear that this is no longer entirely convincing. The numerous advocates of autonomous weapons systems point to their higher capabilities in terms of stress resistance, accuracy or endurance. In other words, the elimination of genuine human weaknesses supposedly leads straight to »clean, ethical and legally impeccable killing«. Their assumption here, however, is that lethal actions by machines, on one hand, and by humans, on the other, as well as lethal mistakes, are ethically equivalent. If one rejects this premise – and there are good reasons for doing so, as we have seen – then one cannot readily gauge

147. US Delegation Opening Statement, CCW Expert Meeting Lethal Autonomous Weapons Systems, Geneva, 13 April 2015, p. 1, [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/8B33A1CDBE80EC60C1257E2800275E56/\\$file/2015-LAWS-MX-USA+bis.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/8B33A1CDBE80EC60C1257E2800275E56/$file/2015-LAWS-MX-USA+bis.pdf) (accessed on 27 April 2015).

the deployment of autonomous weapons in terms of the existing rules, no matter what additional – legal or mechanical – safeguard mechanisms are used. Looked at in this way, the development of machine autonomy represents a real turning point. Recognising this would mean reassessing the problem from the ground up and laying down new, possibly even much higher legal standards than those currently found in international humanitarian law. However, there are currently no indications for such an approach in contemporary state practice. In light of the current course of the discussion, on the occasion of the Geneva experts' meeting – but above all against the background of the current global political situation overall – such a fundamental discussion is not really on the cards at the international level for the foreseeable future.

## 8. Further Proposals

Jeffrey Thurnher proposes the establishment of four control mechanisms that would always have to be taken into account before deploying autonomous weapons systems. First, there must be clearly defined rules of deployment that are always observed. Second, deployment scenarios should be temporally and spatially limited from the outset. Third, autonomous systems should be deployed only in conventional armed conflicts between states, but not in civil wars or to curb insurrections. Fourth, human controls in the form of a possibility of intervention (»on the loop«) must always be guaranteed.<sup>148</sup>

Anderson and Waxman, in turn, propose a multilateral process implemented by experts, similar to the one that led ultimately to the preparation of the (non-legally binding) Tallinn Manual on the applicability of international law to cyber warfare.<sup>149</sup> Such a process should examine how autonomous weapons systems can be adequately regulated by the current rules of international law and work out corresponding (non-binding) guidelines.<sup>150</sup>

148. Jeffrey S. Thurnher, No One at the Controls: Legal Implications of Fully Autonomous Targeting, *Joint Force Quarterly*, Vol. 67, 2012, p. 77, 82 f.

149. Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare*, Cambridge 2013, <https://ccdcoe.org/tallinn-manual.html> (accessed on 17 February 2015).

150. Anderson/Waxman, see above note 65, p. 24.



## VII. Conclusions and Recommendations for Action

### **Autonomous weapons systems: present and future**

- Autonomous weapons systems are not merely a further development of existing weapons systems. In the long run they will change the quality of warfare and thus mark a turning point in military technology.
- At present, there are still no completely autonomous weapons systems. The current debate is oriented towards the future and thus necessarily fraught with uncertainties regarding realistic modes of deployment and strategic advantages.
- What is certain, however, is that far beyond the debate on »combat robots« the relevance of autonomous systems will increase, over the long term, at all (also higher) levels of military and strategic decision-making. This development will not occur all of a sudden, but gradually.

### **International humanitarian law and proposals for better weapons inspection**

- Autonomy and increasing automisation in military systems are not problematic in themselves from an international law standpoint. Autonomously operating minehunters, bomb disposal robots and independently operating ambulance and transport vehicles are examples of rather uncritical fields of application.
- It is beyond question that autonomous weapons systems may be deployed only when they can comply safely with the rules of international humanitarian law, even under combat conditions. Whether, to what extent and when technical development will make this possible – particularly under more complex mission conditions – remains unclear at present; one can only speculate. The more closely defined the (temporal and spatial) scope of deployment and range of tasks of autonomous weapons systems are, the higher the likelihood that such rules can be complied with.
- In this context, many states, at the – so far informal – expert rounds in Geneva in 2014 and 2015, have called for improved implementation of the weapons inspection obligations laid down in Article 36 Additional Protocol I

by means of more transparency, exchange of information and the elaboration of recommendations (»good practices«) for adequate inspection of autonomous weapons systems. This approach is desirable and helpful.

- This would be only one of several steps, however. Non-binding recommendations and improved weapons inspection within the meaning of Article 36 Additional Protocol I alone are insufficient to deal with the complex problems to which autonomous weapons systems give rise. A much more open and broad debate on autonomous (weapons) systems is therefore called for. Germany should lobby for this internationally.
- This is because even if autonomous weapons systems in the future could be programmed in such a way that they operate in accordance with the rules of international (humanitarian) law, and inspection could verify that this is the case, this still does not answer the fundamental question of whether it is ethically and legally justifiable in the first place to delegate decisions on the deployment of (potentially lethal) force to autonomous systems. Combatants and enemy fighters, basically, may be attacked and killed in armed conflict in accordance with international law. Current law, however, says nothing about whether they may be attacked and killed by autonomously operating machines.

### **The requirement of »meaningful human control« as a starting point for future regulatory approaches**

- The starting point of the more extensive debate that is now urgently needed should be the basic assumption that »critical decisions« – in other words, decisions that concern important legal interests, such as the right to life and the right to bodily integrity – may not be delegated to fully autonomous systems on legal (international law) and ethical grounds. In particular, decisions on life and death must always be subject to the ultimate decision and control of a human being.
- This restrictive position is in keeping with the conception of humanity and the guarantee of human dignity of Germany's Basic Law, as well as being the official position of the German government.<sup>151</sup> It is questionable,

151. General Statement by Germany, CCW Expert Meeting Lethal Autonomous Weapons Systems, Geneva, 13–17 April 2015, p. 2, [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/97636DEC6F1CBF56C1257E26005FE337/\\$file/2015-LAWS-MX-Germany.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/97636DEC6F1CBF56C1257E26005FE337/$file/2015-LAWS-MX-Germany.pdf) (accessed on

however, whether and how a corresponding position can be established as legally binding at the international level and with universal validity, as far as possible.

- The requirement of »meaningful human control« over critical decisions has crystallised from the – so far informal – discussion rounds in Geneva as probably the most promising starting point for this initiative.
- Any demand for meaningful human control over critical decisions implicitly rules out complete autonomy in these areas. That means that the discussion of »meaningful human control« concerns a ban on autonomy in certain areas. Going beyond this, more comprehensive prohibition of autonomous (weapons) systems appears unrealistic at the international level in light of the debate so far.
- Because of its openness to interpretation the concept of »meaningful human control« also harbours clear risks, however. But it is also precisely because of this openness that it could prove to make international consensus on the issue possible. It provides an opportunity to increase

the emphasis on ethical considerations and the protection of human dignity in the debate and future regulations even at the international level.

- The challenge now is to frame the concept in sufficient detail. Which decisions are to be regarded as critical and what »meaningful human control« really means require further discussion, also involving technical, military, legal and ethical expertise. It is clear that a genuine possibility of exercising control, also under mission conditions, over the system's critical decisions must always be maintained.
- Against this background, in international bodies such as the United Nations – especially within the framework of the UN Weapons Convention, but also beyond as the case may be – Germany should continue its efforts to ensure that the highest possible level of human control is required in relation to critical decisions and that fully autonomous lethal weapons systems are banned under legally binding international law.

---

27 April 2015). The precise wording is: »We will not accept that the decision to use force, in particular the decision over life and death, is taken solely by an autonomous system without any possibility for a human intervention in the selection and engagement of targets.« This formulation – in particular the words »solely« and »any possibility«, open up room for interpretation and by all means allow gradations with regard to the requisite control.



### About the author

**Robin Geiß** is Professor of International Law and Security at the University of Glasgow. Previously he was Professor of International and European Law at the University of Potsdam. Prior to that, he had worked as Legal Adviser to the International Committee of the Red Cross (ICRC) in Geneva. He is the author of *Failed States* (Duncker & Humblot 2005) and co-author of *Piracy and Armed Robbery at Sea* (OUP 2011).

This publication is the English translation of the initial study *Die völkerrechtliche Dimension autonomer Waffensysteme* from June 2015.

### Imprint

Friedrich-Ebert-Stiftung | International Policy Analysis  
Hiroshimastraße 28 | 10785 Berlin | Germany

Responsible:  
Dr Michael Bröning, Head, International Policy Analysis

Tel.: ++49-30-269-35-7745 | Fax: ++49-30-269-35-9248  
[www.fes.de/ipa](http://www.fes.de/ipa)

To order publications:  
[info.ipa@fes.de](mailto:info.ipa@fes.de)

Commercial use of all media published by the Friedrich-Ebert-Stiftung (FES) is not permitted without the written consent of the FES.

International Policy Analysis (IPA) is the analysis unit of the International Dialogue department of the Friedrich-Ebert-Stiftung. In our publications and studies we address key topics of European and international politics, economics and society. Our aim is the development of policy recommendations and scenarios from a social democratic perspective.

This publication appears within the framework of the working line »European Foreign and Security Policy«, editor: Anna Maria Kellner, [Anna.Kellner@fes.de](mailto:Anna.Kellner@fes.de); editorial assistant: Nora Neye, [Nora.Neye@fes.de](mailto:Nora.Neye@fes.de).

The views expressed in this publication are not necessarily those of the Friedrich-Ebert-Stiftung or of the organization for which the author works.

This publication is printed on paper from sustainable forestry.



ISBN 978-3-95861-271-6