JUSTITIA ET PAX
EUROPA

The Dangers of Lethal Autonomous Weapon Systems

Bernhard Koch (itfh)/Niklas Schörnig (PRIF)

At the end of July 2015, more than 2800 scientists signed an open letter published by the American *Future of Life Institute* calling for a ban on 'autonomous weapon systems' which were beyond 'meaningful human control'. Within a very short space of time, the call had attracted more than 16,000 supporters, including such famous names as Stephen Hawking, Elon Musk – the founder of Teslar and SpaceX – and one of the co-founders of Apple, Steve Wozniak.[1] All warn of an arms race which will result in weapon systems which can select and engage targets without any further human control. These systems are generally referred to today as 'autonomous weapon systems' or AWS. However, this is not the first call warning about such systems, which are still in development. As far back as 2013, almost 300 robotics and computer scientists spoke out against these types of system[2] and, as early as 2010, a network of concerned scientists came together to found the *International Committee for Robot Arms Control* (ICRAC), in an effort to draw attention to the dangers of autonomous weapon systems.[3] For many scientists, autonomous weapon systems represent one of the biggest threats to our peaceful co-existence in the future.

To enable us to comprehend the scientists' concerns, we first need to understand what an autonomous weapon system is. At present, many military systems already enjoy a high level of automation/'semi-autonomy'. Currently-unmanned fighter planes, often referred to as fighter drones or *Unmanned Combat Aerial Vehicles* (UCAVs), are still operated remotely by a pilot. However, they can already execute many tasks automatically/semi-autonomously. This means that many military drones which are currently in deployment can, for example, fly off by themselves, without any human assistance, along paths indicated by GPS coordinates, and also take off and land with virtually no human intervention. An unmanned test plane [*demonstrator*], the American X-47B, even managed to land on an aircraft carrier in Summer 2013 without human help[4] – a manoeuvre that commands respect, even from experienced fighter pilots. Land-based systems, such as the South Korean sentry robot SGR-A1[5] or the Israeli Guardium-System[6], are able to patrol predetermined areas and report abnormalities or intruders to headquarters. These capabilities are mirrored in the civilian field, where, for example, the first 'autonomous' cars are merging with ordinary traffic in an entirely successful fashion.

Even if the precise definition of the term 'autonomous weapon system' is still subject to heated debate in specialist circles, there are nevertheless a few core characteristics on which most experts agree. As yet, the most concise definition can be found in Directive 3000.09, drafted by the American Department of Defense and dated 21 November 2012. It describes the term 'autonomous weapon system' as follows:

> "A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation".[7]

Two key aspects of this definition are significant: first, that targets are selected and engaged by a system without further human input. Second, it makes no difference when it comes to the definition whether a human being is at least involved in a supervisory capacity and can halt the system in the event of a clear error, or whether even this possibility is out of his/her hands. With autonomous weapon systems, the human being is alloted the role of supervisor in the best-case scenario, and in the worst-case scenario, (s)he becomes a spectator who must rely completely on the system's computer algorithms without being able to intervene.

A glance at current military systems reveals that many states with highly-developed military technology already possess so-called 'assistance systems', which are decisive when it comes to supporting people's decisions as to which targets to engage as a matter of priority.[8] To give an example, the pilot of a modern fighter jet can scarcely make out the enemy any more. Computers register and classify possible adversaries at a distance, far beyond the pilot's visual perception. The decision to go on to fire a weapon may still lie with the human pilot, but s/he has to rely entirely on his computer if s/he wants to avoid mistaking a passenger plane for a fighter plane. The pilots of the drones which are currently available must also activate the weapons themselves. But the move to transfer this final decision from man to machine seems but a heartbeat away. Autonomous weapon systems have already been developed for certain scenarios and are even in deployment. Some self-defence systems which aim to protect ships or territories/encampments ('Phalanx'[9], 'Iron Dome'[10], 'Mantis'[11]) are able to record and classify approaching missiles or rockets and engage them

1 cf. *http://futureoflife.org/AI/open_letter_autonomous_weapons*; retrieved 18/8/2015.
2 cf. *http://icrac.net/2013/10/computing-experts-from-37-countries-call-for-ban-on-killer-robots/*; retrieved 18/8/2015.
3 cf. *www.stopkillerrobots.org/chronology/*; retrieved 18/8/2015.
4 cf. *www.navy.mil/submit/display.asp?story_id=75298*; retrieved 15/8/2015.
5 cf. *www.dailytech.com/GuntotingSentryRobotsDeployedInSouthKorea/article19050.htm*; retrieved 15/8/2015.
6 cf. *www.timesofisrael.com/as-google-dreams-of-driverless-cars-idf-deploys-them/*; retreived 17/8/2015.

7 cf. *www.dtic.mil/whs/directives/corres/pdf/300009p.pdf*, S. 13f; retrieved 18/8/2015.
8 cf. For example Schörnig 2014.
9 cf. *www.raytheon.com/capabilities/products/phalanx/*; retrieved 18/8/2015.
10 cf. *www.raytheon.com/capabilities/products/irondome/*; retrieved 18/8/2015.
11 cf. *www.army-technology.com/projects/mantis*; retrieved 18/8/2015.

autonomously within a matter of split seconds. This said, the systems are still mostly operated in such a way that it is still down to a human being to confirm the use of weapons. However, the technology for a fully-autonomous mission is ready. Other systems, such as the Israeli Harpy Drone, fly in circles above an area until such time as they are registered by an enemy radar beam. They then destroy the enemy radar without any further human clearance.[12]

The systems available to date are admittedly not (yet) aimed at people. Systems which select people as targets and kill these people are generally known internationally as '*Lethal Autonomous Weapon Systems*' ('LAWS'). As this paper goes to press (2015) these systems are not yet in use. However, it would not require much technology to develop autonomous weapons which also target and kill people. This would mean for example that it would be conceivable to arm the aforementioned patrol system with a weapon and have it shoot at every registered intruder. One could also imagine future systems which kill all people in a given area whom they are unable to identify from a photo database, or killing machines which would be given a specific picture, set off in pursuit of the person to be killed and not rest until they had completed their mission.

Those in favour of highly-automated systems often argue that even autonomous systems would always remain under human control. However, this assessment is not particularly compatible with military rationale, given that military logic ultimately calls for the capabilities of a system to be fully exploited – if only because there is a fear that the opponent could do likewise and one could put oneself at a disadvantage by placing limits on oneself. After all, the fear is that a state which does away with human decisions could gain a decisive split second in critical missions, thus emerging victorious from a conflict situation. As in a classic arms race, what we can expect to see is the human factor disappearing because it is the slowest link in the chain – something which is at odds with the intentions declared at the outset.

Autonomous weapons and international law

So roboticists and computer scientists are therefore not asking whether it will be possible to develop such autonomous and lethal weapons. Instead, what concerns them is how reliably these weapons can be deployed in line with the legal guidelines for the deployment of weapons. Compatibility with International Humanitarian Law or IHL, whose most important texts are laid out in the 1949 Geneva Convention (GC) and the first two additional protocols (AP) added in 1977, is of real siginificance here. International humanitarian law recognises – roughly speaking – two core principles, to which all parties involved in an armed conflict[13] must adhere. The first is the distinction requirement, i.e. the fact that one must be able to distinguish between legitimate targets ('combatants'; AP I, Art. 43) and protected people ('Civilian Population'[14]; AP I, Art. 50/51; AP II, Art. 13). While it is permitted to attack the opposing side in an armed conflict, civilians must be spared as a matter of principle. The second core criterion concerns the proportionality of means: in a combat situation, only those means may be used which are appropriate to the military purpose.[15] Damage caused to civilians and civilian objects must never be excessive. Nevertheless, the two principles combined do not guarantee the absolute protection of civilians. If civilians are killed in a military operation in spite of precautionary measures having been taken, this is admissible, provided that the civilians in question were not the actual target of the attack and the expected military 'advantage'

was high enough (AP I, Art. 51, 5 b). The responsibility for carrying out an attack in which the possibility of civilian casualties cannot be ruled out then falls to the commanding officer, who may have to answer for him/herself in a court of law at a later date.

This is where the critics of autonomous systems come in. There is a broad consensus amongst experts on international law that the deployment of autonomous weapons systems is also bound by the rules of international humanitarian law.[16] So the debate raises the question as to whether these 'self-deciding' systems are in any way able to comply with the law. If not, they ought not to be used in an armed conflict. But, given the current state of technology, could one conceive of a situation in which autonomous weapon systems were able to implement international law? On this, some roboticists are doubtful as to whether significantly more complex computer systems, of the kind we are familiar with today, will be able to clearly differentiate between civilians and combatants.[17] The asymetrical nature of present-day conflicts is precisely what makes it extraordinarily difficult, even for human soldiers, to tell every time who is a dangerous adversary and who is a harmless civilian. According to these experts, robots will also find this a real challenge in the future. Hence the experts' concern that imprecise systems will lead to significantly higher civilian victims. The question remains as to whether computers will ever be able to weigh up the proportionality of the means that have been deployed. As to whether an attack justifies civilian 'collateral damage' or not, that is where tactical and strategic issues come into play. It is true that humans can also make mistakes when it comes to these kinds of decision – the history of war has plenty of examples of such blunders –, but the fact remains that a human being cannot get around assuming, for the most part legal, but at best moral responsibility. To sum up: there is good reason to argue that the complexity of the task at hand casts doubt upon whether it is possible to 'programme' international humanitarian law into a robot.

Nevertheless, there are also roboticists who, while allowing that it is not possible for robots to implement international law in a perfect fashion, put the case that it is at least possible to teach computers and/or lethal autonomous weapon systems the legal requirements, at least to such a level that would mean that, on average, they did not make worse decisions than human decision makers. For example, roboticist Ronald Arkin is at present developing a form of software which he calls 'ethical governor'. Systems based on this software would, according to Arkin, avoid many human errors on the battlefield which are prompted by emotional reactions, for example acts of revenge because a comrade had just been killed or firing too hastily because a soldier wrongly suspected that s/he was under threat. In Arkin's view, such systems would possibly act in a 'more humane' way than humans on the battlefield.[18] However, even if it was actually possible to devise this kind of software, key questions would remain unanswered, the most obvious concerning whether there are any incentives to turn off the 'ethical software' in certain situations, e.g. where this would create a military advantage, inter alia due to the system reacting more quickly. The software could also be deactivated by self-serving dictators in order to use the systems against their own people. It also remains to be seen how safe lethal autonomous weapons are from being tampered with by hackers. Since many of the components used are 'off the shelf' products in an effort to keep costs down, the required level of security can only be obtained at great financial cost, if at all. Finally, we are faced with the question of human control. Advocates of highly-automated systems may argue that autonomous systems would always remain under human control. In their view, the systems merely carry out people's orders and could also be deactivated in the event of a clear mistake. But, given the complexity of autonomous systems, we cannot be sure whether these machines

---

12 cf. *http://defense-update.com/directory/harpy.htm*; retrieved 18/8/2015.

14 In present-day conflicts, 'Civilians (who) take a direct part in hostilities' (AP I, Art. 51, 3; AP II, Art. 13,3; GC comm. Art. 3) have a significant role to play. They do not come under the category of protected people.

15 Recent times have seen the International Committee of the Red Cross (ICRC) in particular rightly stressing once again that unnecessary violence is also inadmissable when carried out against opposing combatants (cf. For example International Committee of the Red Cross 2013, 4, with reference to the St Petersburg Declaration of 1868).

---

16 cf. for example Geiß 2015.

17 cf. Human Rights Watch 2012.

18 cf. Arkin 2009.

really will always implement the instructions in the manner in which the human commander imagined they would. After all, an inherent part of highly complex systems is the reality that not every system status can be thought through or even tested in advance. One can thus conceive of situations in which an unusual combination of variables leads to a system performance which could not have been foreseen – with potentially fatal consequences. This applies in particular when autonomous weapons converge and the software from one weapon 'interacts' with that of the other – one with which it is not familiar. Whether the human still has time left to intervene is questionable. From a purely military perspective, it will often make more sense in future *not*[19] to equip autonomous systems with active communications and an option for intervention, precisely in order to rule out from the word go external attempts to tamper with the technology. However, if humans surrender the possibility of being able to control the technology because they are afraid it will be tampered with, they are putting themselves totally in the hands of highly-complex computer algorithms which, in most cases, they cannot understand themselves without the aid of a computer.

The ethical dimension

The fundamental reason behind technical instruments being developed is to carry out specific actions with them and achieve specific goals. Any ethical evaluation of an object must therefore first be geared towards the types of action which are carried out with this object `1a`, and the objectives for which the instrument is intended `1b`. However, there is also a need to consider possible risks arising either from the very existence of such an instrument or object or its application. Third, any ethical evaluation must take care to ensure that technical progress does not put paid to our capacity to make ethical judgments per se. This can happen for example when there is absolutely no way to ascertain who was behind the action taken, so that it is also no longer possible to apportion blame in a plausible way. It can also arise when the thinking demonstrated by the technical device and expressed through its instructions is the kind of thinking that eliminates ethical terms and the language of morality as such. The three aspects all fuel the great sense of fear regarding how ethical it is to develop military robotics; since the problems to be discussed affecting the three domains are very far from being solved, a provisional ban on any development of 'lethal autonomous weapon systems' is advised as a matter of urgency.

We have been asked to keep to a word limit, so the following is a rough list of individual aspects to be taken into consideration:[20]

`1a` Autonomous weapon systems are developed in order to carry out violent acts. Although violence can be justified in a very limited set of circumstances[21], for example in places where it attempts to ward off illegitimate attacks, even this kind of counter-violence must always be in proportion to the assets under threat. Deadly violence towards humans can therefore, generally speaking, only be justified where people are fearful for their own lives due to illegitimate violence and retaliate with violence towards those who pose a threat to their own lives. Deploying military robotics leads us into the paradoxical situation whereby violent actions against the robots do not in themselves justify any deadly counter-violence against a human attacker. Even pre-emptive attacks carried out in the absence of a direct threat, e.g. in so-called 'targeted killings'

or 'signature attacks', are hardly justifiable from an ethical point of view.[22]
`1b` The goal of violent action can only be to overcome violence and achieve a state of peace for human beings. Control maintained by robotics does not meet the requirement of peace for human beings. It smacks of pure power play and does not leave people free to perform good deeds but rather suppresses them in a kind of conditioning. Violence aided and carried out by robots does not offer any lasting prospect of peace.

`2` The risks inherent in 'autonomous weapon systems' are incalculable. The aforementioned interaction problems with autonomous systems, which may, for example, arise when ones own autonomous robots interact with other autonomous robots, cannot be predicted at this moment in time, probably also due to the absolutely fundamental characteristics of these systems. Even if it is thought that the probability of such 'emergent phenomena' is small (particularly when compared with possible users of these systems and the likelihood of that) a kind of *precautionary principle* ought to be reason enough to avoid stirring up a catastrophe, the magnitude of which we can only guess at.

`3` Autonomous weapon systems stand accused of occasionally muddying the waters when it comes to ascribing the specific outcome of a weapon to the author responsible. They sometimes even make this impossible. Such concerns are indeed justified – possibly not in every individual case where this type of system is used, but the further away the impact of the decision to deploy is felt, and the more abstract the targets provided to the machine in advance, the more complicated it becomes to determine who is behind the concrete action. However, when thinking through the ethics of a situation, it is crucial that we know who authored the action. Even if we wished to pass an ethical judgment on structures – for example, commanding or programming structures –, we would need to conclude by evaluating them according to the type of action taken, an action brought to life in these structures. It is true that positive law dictates often enough who is reponsible and to what extent, but such pure legislation loses its ethical power when it is applied in an entirely arbitrary fashion.[23] The ultimate course of action resides in the genuinely human practice (a system which cannot be cheated) of holding people to account and accounting for ones own actions.[24]

The language which we use to refer to the implementation of autonomous robotics is also dangerous. When we say that the robots 'would be making a decision themselves', we can only state that metaphorically, as an anthropomorphic analogy. This can be justified for as long as people are aware that we are speaking metaphorically. There is a danger that soon we shall no longer be using human decision as understood by the living world as our point of reference when it comes to when to apply this expression but rather the machines' 'decision', and we may also be interpreting and evaluating people's decisions on this basis. This would lead to the crucial *Humanum* inherent in humans carrying out actions for themselves falling completely by the wayside, with the consequence that the actions of soldiers would be measured according to how they were enforced by robots. However, since we can only judge a robot by its appearance, e.g. its results, this purely results-oriented reference would remain the sole parameter of evaluation for military action – and thus safeguard comparability. But in so doing, we fail completely to do justice to the ethical requirement which we place on our action as *behaviour*.

---

19 Translator's italics
20 The factors laid out here have been presented in a very rudimentary fashion and need to be differentiated in a far more thoroughgoing manner. In the past, emphasis was placed on this issue in the field of analytical ethics, which led to a markedly more precise debate, though it did not manage to solve the problem once and for all.
21 The possibility that violence may be justified does not however necessarily exclude the voluntary renunciation of violence on ethical grounds.

---

22 This already applies to remote-controlled military robotics such as armed drones (UCAVs). On 'targeted killings', cf. for example Koch 2015.
23 For example, the question is sometimes asked as to whether the individual soldier or his/her commander or possibly the programmer of an autonomous system should bear responsibility for a malfunction. The law can be of assistance here, but not without any kind of evidence in as to prerogatives. When there are great distances between the programmer's act or the act of input, and the process that demonstrates the problematic result, such evidence is completely diluted.
24 cf. Koch 2014.

Closing remarks

Autonomous lethal weapon systems represent one of the greatest challenges to the peaceful co-existence of peoples. They are not just highly problematic from the point of view of international law. Rather, they also threaten to have a destabilising impact on security policy. However, the most significant thing to bear in mind is that, in almost all cases, using them is ethically unacceptable. Unless we move to ban the development of lethal autonomous weapons as soon as possible, it is only a matter of time before they are developed and procured. In the end, human beings would no longer be the decision-makers, because the kind of classic military logic that drives the arms race comes into effect here, giving the lie to supporters who claim that, even with highly complex systems, the desire is for humans to remain the ultimate decision-makers. It is encouraging that the international community of nation states has now also come out in recognition of the problem of LAWS. As far back as 2013,

Christof Heyns, UN Special Rapporteur on extrajudicial, summary or arbitrary executions, published a critical report on LAWS and called on the community of states to halt the development of such systems in order to find out more about the dangers that they pose (Heyns 2013). In 2014 and 2015, expert talks were held within the framework of the UN Weapons Convention, known as the *Convention on Certain Conventional Weapons* (CCW), held in Geneva. So there have been enough warnings so far. Now it is time for the international community also to nail its colours to the mast as we draw closer to decisions on CCW, and it should press for a ban on autonomous lethal weapons. The goal must be, as various NGOs have suggested,[25] to retain 'meaningful human control' over" all weapon systems. We surely cannot go down the path which ends with computer algorithms making life or death decisions.

---

25 **cf**. www.article36.org/weapons-review/autonomous-weapons-meaningful-human-control-and-the-ccw/; retrieved 20/9/2015.

30/09/2015

Authors

Dr. Bernhard Koch, Deputy Director, Institute for Theology and Peace, Hamburg (*http://www.ithf.de*)

Dr. Niklas Schörnig, Senior Researcher, Leibniz-Institut HSFK (Hessische Stiftung Friedens- und Konfliktforschung/Hessian Foundation for Peace and Conflict Studies), Peace Research Institute Frankfurt (PRIF) (*http://www.hsfk.de*)

Cited Literature and Further Reading

Abney, Keith 2013: Autonomous robots and the future of just war theory, in F. Allhoff, N. G. Evans and A. Henschke (ed.): Routledge Handbook of Ethics and War. Just war theory in the twenty-first century, New York/London: Routledge, 338-51.

Arkin, Ron. 2009. Governing Lethal Behavior in Autonomous Robots. Boca Raton: CRC Press.

Dinstein, Yoram 2011: War, Aggression, and Self-Defence, 5th edition, Cambridge: Cambridge University Press.

Geiß, Robin 2015: Die völkerrechtliche Dimension autonomer Waffensysteme, Berlin: Friedrich Ebert Stiftung.

Geneva Academy of International Humanitarian Law and Human Rights 2014: Autonomous Weapon Systems under International Law, academy Briefing No. 8, November 2014 (online at: *www.geneva-academy. ch/docs/publications/Briefings%20 and%20In%20breifs/Autonomous%20 Weapon%20Systems%20under%20 International%20Law_Academy%20 Briefing%20No%208.pdf*; retrieved 26/9/2015)

Heyns, Christof 2013: Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, A/HRC/23/47, New York: United Nations Human rights Council (online at: *www.ohchr.org/Documents/HRBodies/ HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf*; retrieved 29/9/2015).

Human Rights Watch 2012: Losing Humanity. The Case against Killer Robots, Washington, DC: Human Rights Watch (online at: *https:// www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots*; retrieved 27/9/2015).

International Committee of the Red Cross 2013: Handbook on International Rules Governing Military Operations, Geneva: ICRC (online at: *https://www.icrc.org/eng/assets/files/ publications/icrc-002-0431.pdf*; retrieved 29/9/2015)

International Committee of the Red Cross 2014: Expert Meeting Autonomous Weapon Systems. Technical, Military, Legal and Humanitarian Aspects, Geneva, Switzerland, 26 to 28 March 2014, Report, Geneva: ICRC (online at: *https://www.icrc. org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014*; retrieved 27/9/2015).

Kershnar Stephen 2013: Autonomous weapons pose no moral problems, in: B. J. Strawser (ed.): Killing by Remote Control. The Ethics of an unmanned military, Oxford: Oxford University Press, 229-245.

Koch, Bernhard 2014: Zur ethischen Bedeutung von Zurechenbarkeit, in: M. Gillner/V. Stümke (Hrsg.): Kollateralopfer. Die Tötung von Unschuldigen als rechtliches und moralisches Problem, Baden-Baden: Nomos & Münster: Aschendorff Verlag, 113-137.

Koch, Bernhard 2015: Targeted Killing. Grundzüge der moralphilosophischen Debatte in der Gegenwart. In: V. Bock/J. J. Frühbauer/A. Küppers/C. Sturm (Hrsg.): Christliche Friedensethik vor den Herausforderungen des 21. Jahrhunderts, Baden-Baden: Nomos & Münster: Aschendorff Verlag, 191-206.

Krishnan, Armin. 2009. Killer Robots. Legality and Ethicality of Autonomous Weapons. Farnham: Ashgate.

Leveringhaus, Alexander 2016: Ethics and autonomous weapons: technology and armed conflict in the 21st century (forthcoming)

Purves, Duncan/Jenkins, Ryan/Strawser, Bradley J. 2016: Autonomous Machines, Moral Judgement, and Acting for the Right Reasons. In: Ethical Theory and Moral Practice (forthcoming)

Sassóli, Marco 2014: Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified. In: International Law Studies 90, 308-340 (online unter: *https://www. usnwc.edu/getattachment/96b691c2-d425-47d7-b6c9-1c1bd691d01d/ Autonomous-Weapons-and-International-Humanitarian-.aspx*; retrieved 29.9.2015).

Schmitt, Michael N. / Turnher, Jeffrey S. 2013: "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict, in: Harvard National Security Journal 4: 2, 231-281.

Schörnig, Niklas 2014: Automatisierte Kriegsführung - Wie viel Entscheidungsraum bleibt dem Menschen?, in: Aus Politik und Zeitgeschichte 64: 35-37, 27-34.

Singer, Peter 2009: Wired for War. New York: Penguin.

Sparrow, Rob. 2007: Killer Robots, in: Journal of Applied Philosophy 24:1, 62-77.

UNIDIR 2014: Framing Discussions on the Weaponization of Increasingly Autonomous Technologies, UNIDIR: *www.unidir.org/files/ publications/pdfs/framing-discussions-on-the-weaponization-of-increasingly-autonomous-technologies-en-606.pdf*; retrieved 24/6/2015.

Williams, Huw. 2013: The next step: advancing from unmanned to autonomous, in: Jane's International Defence Review, October 2013, 62-65.