

Müller, Vincent C. (2016), ‘Autonomous killer robots are probably good news’, in Ezio Di Nucci and Filippo Santoni de Sio (eds.), *Drones and responsibility: Legal, philosophical and socio-technical perspectives on the use of remotely controlled weapons* (London: Ashgate), 67-81. <http://www.ashgate.com/isbn/9781472456724> DOI: 10.4324/9781315578187-4

Autonomous Killer Robots Are Probably Good News*

Vincent C. Müller

ACT/Anatolia College, Thessaloniki

<http://www.sophia.de>

Abstract: Will future lethal autonomous weapon systems (LAWS), or ‘killer robots’, be a threat to humanity? The European Parliament has called for a moratorium or ban of LAWS; the ‘Contracting Parties to the Geneva Convention at the United Nations’ are presently discussing such a ban, which is supported by the great majority of writers and campaigners on the issue. However, the main arguments in favour of a ban are unsound. LAWS do not support extrajudicial killings, they do not take responsibility away from humans; in fact they increase the ability to hold humans accountable for war crimes. Using LAWS in war would probably reduce human suffering overall. Finally, the availability of LAWS would probably not increase the probability of war or other lethal conflict—especially as compared to extant remote-controlled weapons. The widespread fear of killer robots is unfounded: They are probably good news.

Keywords: killer robot, LAWS, LAR, AWS, drone, robotic weapon, automated warfare, responsibility, distinction, proportionality, just war, humanitarian law, Geneva convention, regulation, utility

1. Definition

The autonomous robotic systems that are the subject of this paper would be able to select and attack targets without intervention by a human operator. While the initial command to attack would be given by a human, the robot then has a degree of autonomous ‘choice’ for its actions.

Some examples of systems that fall under the scope of this paper are automated radar-guided gun systems to defend ships that have been in use since the 1970ies (e.g.

* The policy recommendations of this paper are spelled out in (Müller and Simpson 2014); the philosophical points about responsibility are discussed in detail in (Simpson and Müller forthcoming). – I am grateful for the cooperation with Tom Simpson. My thanks for comments on earlier drafts go to: Stuart Armstrong, the editors of this volume, an anonymous reviewer, audiences in Delft, Athens, Aarhus and at the online ShanghAI Lectures.

the US ‘Phalanx’ Close-In Weapon System). Current versions can autonomously identify and attack oncoming missiles, rockets, artillery fire, aircraft and surface vessels according to criteria set by the human operator. Similar systems exist for tanks, e.g. the Russian ‘Drozd’ (1977-82) and now ‘Arena’ or the German ‘AWiSS’/‘AVePS’ (Active Vehicle Protection System) by Diehl, which has a reaction time below 400ms. The main reason for not having a ‘human in the loop’ in these systems is sheer speed.

Systems with a higher degree of autonomy would include drones, e.g.: “The United Kingdom Taranis jet-propelled combat drone prototype can autonomously search, identify and locate enemies but can only engage with a target when authorized by mission command. It can also defend itself against enemy aircraft” (Heyns 2013, §45). The US X-47B drone can take off and land on aircraft carriers (demonstrated in 2014); it is set to be developed into an ‘Unmanned Carrier-Launched Surveillance and Strike’ (UCLASS) system. It is now quite conceivable that an autonomous drone, perhaps with the size and appearance of a bird, could be commanded to locate (e.g. using cell phone signal), pursue and kill an individual person—rather like a ‘hit man’. In a few years, some states will have the ability to deploy such a killer drone anywhere in the world, for example against someone they consider a ‘terrorist’.

While autonomous air systems are fairly advanced, these kinds of systems will shortly also be available on land, on water and under water, as well as in space. For example, the US Navy has developed and tested small ‘swarm boats’ that can accompany a ship, and protect it from small vessel attacks by detecting and swarming around such vessels—the current version has a ‘human in the loop’ for weapons fire (Smalley 2014).

1.1. Problem

Robotics in warfare will be a major change, comparable to the introduction of planes or perhaps even nuclear weapons (Singer 2009b: 179, 203). One of the questions is whether the use of robotics, especially highly autonomous robotics, constitutes just one major step in the arms race, or whether it is a step that introduces qualitatively new ethical concerns. Many authors and organisations have claimed that killer robots are a serious threat to humanity and should be banned, while others have said there is nothing new here (Arkin 2009, 2013). As the UN Rapporteur says in his careful and detailed report: “Some argue that robots could never meet the requirements of international humanitarian law (IHL) or international human rights law (IHRL), and that, even if they could, as a matter of principle robots should not be granted the power to decide who should live and die. These critics call for a blanket ban on their development, production and use” (Heyns 2013, §31).

I agree that LAWS pose new problems and therefore recommend that they must meet special *technical standards* of reliability, esp. with respect to ‘distinction’ and ‘pro-

portionality’, and that their manufacturers and users abide by *legal standards* that allow allocation of responsibility.

In this policy paper, I provide a general recommendation on the issue whether killer robots should be banned, concluding that they should not. I do so by providing a concise survey of the relevant moral concerns.

1.2. Terminological note

The UN now uses ‘LAWS’, for ‘Lethal Autonomous Weapon Systems’ (Simon-Michel 2014), and I follow this, despite its unfortunate positive connotations. The weapons concerned are also often called ‘LARs’ (lethal autonomous robots; e.g. (Heyns 2013); simply ‘drones’ (European Parliament 2014); ‘killer robots’ (www.stopkillerrobots.org); ‘robotic weapons’ (Leveringhaus and Giacca forthcoming 2014); or ‘unmanned systems’. I think that the systems concerned are not just lethal in their effects but made to be lethal, i.e. they are weapons, so ‘LARs’ is insufficiently precise. ‘Drones’ is too narrow, since we are not only talking about flying systems, and too broad, since present drones are remote-controlled (‘remote-piloted’). ‘Unmanned’ is sexist, but in any case the distinguishing feature here is not whether the human in control is actually inside the system (e.g. the plane) or controlling it from a remote location (as with current ‘drones’), but whether the system has a degree of autonomy. For this reason ‘autonomous weapon system’ (AWS) could be an option, but it is not used. ‘Killer robots’ is apt in that it implies autonomy and is not limited to air systems; while perhaps overdramatic, it makes explicit the moral issue at stake. So I use it together with ‘LAWS’, acknowledging the limitations of both terms.

1.3. Simple slogans

A discussion about weapons, killing, suffering and war often generates heated exchanges and reduction to simple slogans. Slogans may be useful campaigning tools, but do not address the deep moral issues. To forestall misunderstanding, some immediate qualifications are in order. Despite the provocative title, I agree that killing and wars are a great evil. More weapons are generally bad, too; they increase the probability of killing and they are usually an inefficient use of resources. I am also deeply concerned about the current use of drones for extrajudicial killings (see Melone and Koch, this volume, and (Enemark 2014)), and about the importance of civilian gun control (Müller 2015).

1.4. Structure of the paper

After an abbreviated look at the current technical and policy situation, I discuss the four main arguments in this debate: whether LAWS are *inherently* wrong because a) they violate humanitarian law, or b) they make it harder or impossible to assign responsibility for killings in war; and whether *the consequences* of LAWS are good or bad

in the long run—either c) by increasing or decreasing the suffering of war or d) by making war more or less likely. I conclude with a number of policy recommendations.

2. Current situation

2.1. Technological

It is clear that increasingly autonomous weapons are coming. The first systems that make simple ‘attack’ decisions are already in use (see 1.1 above). Remote-controlled air systems have been used extensively, especially in the asymmetric Pakistan ‘drone war’. Remote-controlled water, underwater and ground systems are also in use or close to deployment (see e.g. Singer 2009a). Meanwhile, autonomous driving and sailing systems are at a high level of technological readiness, being tested outside the lab (Smalley 2014). The 2013-2038 ‘Unmanned Systems Integrated Road Map’ of the US Department of Defense (US Department of Defense 2013) foresees increasing levels of autonomy in air/land/sea systems in the coming 25 years. The funding for such systems is massive. Taking the US only, the Department of Defense is currently spending ca. \$5 billion US per year on ‘unmanned systems’ (US Department of Defense 2013: 3), as well as an unknown amount of DARPA’s \$3 billion US per year budget. Further US sources cannot be ruled out. While the US is the world leader in military technology with spending on its military ca. 40% of the world’s defence spending (ISS 2014), other countries are certainly pursuing automated technologies, notably Russia and China.

The enabling technology for autonomous AI is developing apace. The median estimate of probability moving over 50% for high-level machine intelligence with full human abilities is 2040, according to a recent survey of expert opinion (Müller and Bostrom 2016). Even if these estimates turn out to be excessively optimistic, significant autonomy levels in target selection and attack will clearly be possible in the next decade already.

Killer robots are attractive to the military, and thus political funding, for a number of reasons: They reduce the risk to one’s own soldiers, reducing the human and political costs of war. They can be cheaper than human soldiers in the long run, not needing a salary, pension, housing, food or hospitals, etc. They can also outperform humans and human-controlled systems, especially in terms of speed, accuracy and ability to function without rest. They can function in environments where human remote-control is not an option (e.g. under water).

2.2. Policy

Some states, notably the USA, have developed initial policies for LAWS that include a moratorium on systems that do not have a ‘human in the loop’. However, these policies can be changed any time, at the discretion of these states. Indeed, this possibility

is explicitly stated in the US policy: “Autonomous ... weapons systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.” (US Department of Defense 2012: 2)

Current armed unmanned systems deploy lethal force only in a fully human-operated context for engagement decisions. ... As technology advances and more automatic features are introduced, DoD will continue to carefully consider the implications of autonomy in weapon systems to ensure safe and reliable operations and minimize the probability and consequences of failures that could lead to unintended engagements. For this reason, DoDD 3000.09, Autonomy in Weapon Systems, mandates a policy review before entering formal development for proposed weapon systems that would use autonomous capabilities in a novel manner. (US Department of Defense 2013: 24)

The UK is working on the matter: “The pace of technological development is accelerating and the UK must establish quickly a clear policy on what will constitute acceptable machine behaviour in future; there is already a significant body of scientific opinion that believes in banning autonomous weapons outright,” (Ministry of Defense 2011).

An independent policy report from the University of Birmingham on “The Security Impact of Drones” has come out in October 2014 and it recommends a ban on LAWS, since they “fail the test of the ‘laws of humanity and the requirements of the public conscience’” and “it is unclear where responsibility would lie for any unlawful actions by weaponised robots would lie”. The report also cites UK politicians saying “the UK’s position on not wishing to develop such weapons is absolutely clear.” (Birmingham Policy Commission 2014: 64, 65).

3. Moral arguments

The arguments against killer robots fall into two broad categories: principled ones concerning rights and responsibility, and utility considerations deriving from the likely consequences of their use.

3.1. War crimes & international law

Killer robots, like any weapon in war, must comply with the regulations of International Humanitarian Law in the Geneva Convention. If they cannot, they are illegal weapons and their use in war (*in bello*) constitutes a war crime. Their use outside a war is a crime in any case (details in Emmerson 2014). There are two pertinent requirements. They must comply with the *principle of distinction*, i.e. have the ability to discriminate combatants from non-combatants. The legal use of LAWS would thus require a positive identification of enemy soldiers, tanks, airplanes, etc. With present technology, this will be possible only in some situations. In particular, building LAWS’ capable of

identifying major weapons platforms such as tanks or submarines as military targets is likely to be relatively feasible, while building those capable of identifying individual humans as either combatant or non-combatant is likely to be very difficult for some time.

The other pertinent Humanitarian Law principle is the *principle of proportionality*, which requires that damage to civilians is proportional to the military aim. Again, current technology for robots can make some defensible judgments on this issue. There is an important exception to this, however, notably in cases where a system could confirm that *no* collateral damage was likely. In practice, judgments of proportionality could be made by the commander, in the process by which they specify the domain within which a killer robot would operate. A current example is the process of authorisation of the ship defence system ‘Phalanx’. The relevant commander sets the system to ‘weapons free’ only in conditions where there is near-certainty that fast-flying missile-like objects are hostile and therefore legitimate targets. The same process applies for more potent weapons systems. A descendent technology of the Predator could be given an area of ocean which a commander has good reason to believe will be occupied only by the enemy’s navy, and given ‘weapons free’ to attack targets within that. The proportionality requirement would thus be met.

The United Nations has recently urged member countries to “ensure that any measures taken or means employed to counter terrorism, including the use of remotely piloted aircraft, comply with their obligations under international law, including the Charter of the United Nations, human rights law and international humanitarian law, in particular the principles of distinction and proportionality;” and asserted “... the urgent and imperative need to seek agreement among Member States on legal questions pertaining to remotely piloted aircraft operations;” (United Nations 2013, 6s, 17). The European Parliament is more forthright and says “drone strikes outside a declared war by a state on the territory of another state without the consent of the latter or of the UN Security Council constitute a violation of international law and of the territorial integrity and sovereignty of that country;” (European Parliament 2014, principle E). Some, notably the US President, have disagreed, and claim wider rights to self-defence (cf. Schmitt 2014).

Given this context, it is unfortunate that the UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, reported on LAWS at the same time. The normative issues raised by cross-border targeted killings are one thing, those by the use of LAWS are quite another. In addressing the question whether a killing is extrajudicial, it is irrelevant whether the weapons concerned are used directly, controlled remotely, or function autonomously. In any case, Heyns, who calls for a moratorium, admits that “While it is not clear at present how LARs [lethal autonomous robots] could be capable of satisfying IHL [International Humanitarian Law] and IHRL [International Human Rights Law] requirements in many respects, it is

foreseeable that they could comply under certain circumstances, especially if used alongside human soldiers.” (Heyns 2013: 109)

This discussion is clouded by a simple dichotomy of autonomous vs. non-autonomous systems. To the contrary, autonomy is a matter of degree and is relational, i.e. something is autonomous with respect to something else, to some degree (cf. Müller 2012). The vague nature of autonomy has the awkward implication for proponents of a ban that it would be very hard to formulate clear controllable criteria for establishing which systems count as killer robots (cf. Anderson and Waxman 2013).

There is no deep mystery here. The legal situation is clear enough. Weapons that violate the laws of distinction and proportionality in war are already prohibited. Just like any other weapon, killer robots should be used only if doing so is in compliance with the law; otherwise their use is a war crime. But there is no principled reason to suppose that killer robots could not conform to the requirements of distinction and proportionality. Whether or not they will do so is an empirical question. Contrary to popular belief (e.g. Garcia 2014; Sharkey 2008a; Sharkey 2008b, 2012), this is *not* an argument to ban killer robots.

It would be an argument only if autonomous killer robots had some feature that would make the prosecution of illegal killings difficult or impossible. Some have suggested that this is the case: “Nobody is responsible for the killing”, they say. This objection deserves more detailed analysis, to which we now turn.

3.2. Responsibility

3.2.1. The Responsibility Gap

The application of criminal law requires that someone be in the dock. This requirement, along with a platitude about the relation between law and morality, generates a serious problem for the justiciability of LAWS. This section sets out the problem. The next summarises how to resolve it.

For the most part, law ‘piggy-backs’ on morality. That is, many legal ascriptions of responsibility do and should track moral responsibilities which are in an important sense prior to the law. For instance, the law should define legal responsibility for murder in ways that ensure the conviction of only those people who have committed the moral wrong of murder.

The problem for killer robots then arises given the following two premises. First, *prima facie*, it will be possible for people to be very seriously wronged as a result of LAWS’ action. People who should not be killed—such as civilians, or surrendering soldiers, or prisoners of war—may well be. The moral wrong is of such a degree that there ought to be laws that hold as legally responsible those who are morally responsible for those deaths. A war cannot be waged justly if no-one could be morally and legally held accountable for such serious wrongings.

Second, however, it is far from clear who is morally responsible for deaths caused by LAWS. Recall that LAWS are designed to be autonomous; that is, to be able to select targets and make ‘decisions’ to kill people *without* human input. So who is to blame when killer robots get it wrong? It is not clear that the military commander who deployed the system is blame; they were just given the equipment and given the order to use it. Nor is it clear that the robot’s designers are to blame. They were commissioned to make something that is autonomous; it is a sign precisely of the success of their work that the system is able to do so. Finally, nor is it the dumb robot. The ability to take on ethical responsibility and be an appropriate target for reward and punishment is a complicated matter that involves at a minimum having goals and desires, the ability to reflect on these, to act against one’s desires, and to understand the consequences of one’s actions (which is why we usually do not hold small children or animals responsible for their action). It is clear that current systems do not have the necessary properties for responsibility and that, for the foreseeable future, artificial systems will not acquire them—so we do not need a ‘machine ethics’ in that sense (on this issue, see Gunkel and Bryson 2014).

As Rob Sparrow puts it, killer robots threaten to create a ‘responsibility gap’. Yet it is a condition of the possibility of a Just War that such gaps not exist (Sparrow 2007; the concept derives from Matthias 2004). This ‘responsibility ascription problem’ is recognised as a large issue in robotics (on which (see Lichocki et al. 2011); for recent papers in the philosophy in technology (see Battaglia et al. 2014), also Saxon in this volume). As such, killer robots ought not to be used in war. So they ought not to be used at all.

3.2.2. *Living with responsibility gaps*

I deny the thesis that the existence of responsibility gaps morally precludes the use of a technology. The claim is not true in civil life, nor is there reason to think it true in war. Consider the former first. Even in matters of life and death, such as pharmaceuticals or building bridges, we only expect ‘due care’ from those involved (engineers, public inspectors, users, ...) and we specify regulations to ensure that the standards of due care are high. But, if such care is exercised, we do not require that there be no deaths. A certain engineering ‘tolerance’ is accepted, even if we try to keep this as small as practically possible. Within that tolerance we expect the bridge to function correctly, assuming certain environmental conditions. But if these careful estimates were wrong, e.g. we face a stronger earthquake than could be expected, we say that the collapse of the bridge and thus the deaths caused are ‘accidental’: Nobody is responsible. Likewise for drugs. There is net benefit in having effective drugs. Drugs are licenced and prescribed, even though we know that there is a risk that some patients will suffer from serious unintended side-effects of taking them. This is unavoidable due to the biochemical complexity of humans, and the unpredictability for individuals of

the introduction of novel compounds. The benefit of having effective drugs, even with occasional negative side-effects, is shared collectively. It is also enjoyed by individuals *ex ante*, who expect in anticipation to gain benefit from taking the drug, even granting the risk of negative side-effect.

When a system fails within conditions inside those it should successfully tolerate, someone is responsible. In the bridge case, a driver may be responsible for its collapse if she takes a lorry over it that is heavier than the load it was designed and stated to be able to bear. Or it may be the engineer who is responsible, for cutting costs by using inferior materials or just being negligent. It could be the public official responsible for confirming compliance with the appropriate regulations. But, importantly, it is conceptually possible that the bridge collapse and *no-one is responsible*. Suppose the bridge crosses a river, which is usually 2m in depth and in the last 200 years has never had a volume of water exceeding a depth of 5m. But exceptional floods mean that a depth of 8m of water is pressing against the pier, which collapses. It would have been a misuse of resources to build the bridge to withstand that force; if the same margins of error were applied to bridge-building generally, it would be inefficient. In such a case, no-one is responsible for the bridge's collapse. If this is correct, the real problem is not the possibility that things may go wrong in a way that no-one is responsible. Rather, it is how to regulate the tasks of the maker and of the users of the system, in such a way that responsibility is *usually* but not *always* ascribable. Further, it is unclear what reason there would be to make special demands for technologies used in war. There is a responsibility gap, but this is the normal gap that we accept in engineering. (This argument is developed at length in (Simpson and Müller forthcoming).) So, what I need for my argument here is only the thesis that if a technology produces rare cases of killings where no person is responsible, this do not by itself compel us to ban the use of this technology. A strong responsibility principle that allows no responsibility gaps at all is untenable in practice. However, responsibility gaps are clearly undesirable and would rule out systems that systematically or even just frequently produce such cases. I argue that killer robots can be regulated in such a way that they do not produce more responsibility gaps, but rather reduce their frequency: It is already the base that for many killings in war responsibility cannot be allocated.

3.2.3. *Narrowing the responsibility gap*

The responsibility framework outlined above shows how responsibility should be ascribed for many of the wrongful killings that could be committed by killer robots. The technology gives rise to a related and further beneficial effect, which is often not noted. Holding someone accountable for their action, e.g. for actual conviction for a war crime requires reliable information—which is often unavailable. The ability to acquire and store full digital data records of LAWS' action and pre-mission inputs allows a better determination of the facts, and thus of actual *allocation* of responsibility, than is

currently possible in the ‘fog of war’. As well as allowing allocation of responsibility, the recording of events is also likely to diminish the likelihood of wrongful killings. There is already plenty of evidence that, for example, police officers who have to video their own actions are much less likely to commit crimes. So, killer robots would actually reduce rather than widen responsibility gaps.

3.2.4. Regulation and standards

The foregoing has the following implication: moral interest should be focused on the determination of the technical standards of reliability which robots—including killer robots—should meet. The recent EU ‘RoboLaw’ report makes a parallel point, in arguing that we should resist the urge to say that ‘robots are special’ in terms of responsibility. Rather, we should adopt a functional perspective and see whether the new technology really does require new legal regulation, and in which areas (based on Bertolini 2014; Palmerini et al. 2014: 205f). This seems to be a move in the right direction: We already devise automated systems (e.g. automated defence of ships against air attacks) where the ‘rules of engagement’ are put into software. The same ‘due care’ is to be expected for the manufacture and use of LAWS. Just like for civil autonomous cars, we need to specify standards that LAWS manufacturers must abide by. These standards must ensure that the robot acts according to the principles of distinction and proportionality (this is already possible now if one thinks of targeting tanks, ships, planes or artillery, for example). Both manufacturing and distributing LAWS that do not abide by these standard would be a war crime. If a killer robot is manufactured with due care according to these standards but commits a war crime, due to use in situations for which it was not designed or licensed, the crime is the responsibility of the soldier/user. The responsible person for a particular command or action can be identified in the military chain of command – this is a deeply entrenched tradition. Finally, if the soldiers can show that they exercised due care, then the deaths are accidents.

Regulation of LAWS thus requires two components. First, there are the technical standards of reliability which LAWS must meet; pertinently, what degrees of reliability LAWS must meet in terms of distinction and proportionality in their decisions to attack. Second, there are the legal instruments by which accountability is to be exercised over those who fail to manufacture, distribute or deploy LAWS in accordance with those standards. Each dimension—that of technical standards and of law—should be subject to enforcement at the international and national levels.

The proposed policy structure can thus be schematically presented as a matrix:

Legal and technical regulation

	International	National
Legal	International Humanitarian Law	Criminal Law
Technical	Technical standards for performance	Control regimes for technical standards; special standards

3.3. Are killer robots good for us?

Assuming there is no particular ethical problem with humanitarian law or responsibility, the remaining question is whether the consequences of LAWS are good overall, in the long run—do they reduce or increase human well-being? If utility is reduced by the existence and use of these weapons we should not allow them. There are two subordinate questions here. Do LAWS increase or decrease the suffering and loss of war? Do LAWS make war more or less likely?

3.3.1. Reducing the human cost of war

There are a number of points that suggest LAWS would reduce the human cost of war.

- Robots reduce war crimes and crimes in war: they do not rape, do not get angry or afraid, they do not intentionally commit war crimes—unless programmed to do so. They follow orders more closely.
- “One of the great arguments for armed robots is they can fire second,” Joseph W. Dyer, cited in (Markoff 2010)
- Drones are excellent data-collectors, so perpetrators of war crimes are more likely to be caught. This also makes war crimes less likely
- Fewer deaths, injuries and traumas of combatants
- Fewer deaths, injuries and traumas of non-combatants
- Thus less damage to future generations

3.3.2. Making war worse

There are a couple of points that suggest LAWS would make wars worse:

- LAWS have limited judgment and common sense, which will lead to errors and to carrying out orders that violate the law of war.
- Killing is made easier if the command can be passed on to an autonomous system, so proportionality is under threat.

3.3.3. Making war more likely

There are some points that suggest LAWS would make wars more likely:

- With LAWS, a war can be expected to result in less death and injury to the soldiers on the side that has them available (but only slightly, if compared to remote-controlled systems).
- They make wars less bad, generally, and thus wars are more likely to be chosen as a means.
- They make a particular military action easier to decide for a military commander (see Krishnan 2009).
- Fewer losses of soldiers' lives reduce the political hurdle for wars and esp. military action short of war.
- Finally, they make it easier to maintain a 'low level' war for some time, especially if it is an asymmetric war.

3.3.4. *Utility, fairness and arms races*

The reasons why LAWS make a war more likely apply equally to remote-controlled weapons; in fact they apply to any weapon that acts at a distance. Such weapons have always resulted in relative safety for the first users: Goliath probably thought David's sling unfair; the Lateran Council of 1139, Canon 29, banned use of the large cross-bow [ballista] and archery against fellow Christians. The criticism that LAWS lower the risk for attackers and thus make wars and other killings more likely is thus *correct*, but applies to any weapon that the one side has, but the other does not: In other words, it is the result of an on-going arms race. As soon as the other side has acquired the new weapon, the risk of war goes down again.

This is not to say that I think LAWS are part of an unstoppable arms race. Some weapons systems have been banned (anti-personnel mines, chemical weapons) and with nuclear weapons the arms race is highly controlled. We *can* stop developing these weapons—the question is whether it is right to do so, given that they seem to save lives.

Let us note that the reasons why LAWS make wars less bad do not apply to all weapons at a distance. This is so especially for weapons of mass destruction or weapons with poor accuracy—and thus with poor compliance to the humanitarian law requirements of distinction and proportionality.

If killer robots become cheap and easy to obtain or make, then the consequences would certainly be bad – as in any case of a weapon becoming more widely available (see Müller 2015) – so we would do well to prevent this spread.

3.3.5. *Utility Overall*

So, what is the overall utility count? As usual with utility in the long run, this is very hard to say but it seems quite clear that LAWS would do something to reduce the human cost of war. Some reduction is extremely valuable, given how serious the suf-

fering of war is. Further, while LAWS do raise the probability of wars in the short run, there is no reason to think they will do so in the long run. The overall utility calculation depends on the balance of how much less bad wars become and how much more likely they become. How bad the short run raise in probability will turn out depends mainly on which the first parties to acquire them are. Given current military spending, we know who these are: USA, China, Russia. It also depends on how big the difference to remote-controlled systems is, which currently looks minimal. If they do not substantially increase the probability of war, then killer robots are good news for humanity.

4. Conclusions & policy recommendations

I conclude that:

1. Killer robots pose no substantial challenge to humanitarian law
2. Killer robots pose no substantial issue of responsibility
3. Given 1 and 2, the crucial issue is whether the overall consequences of killer robots are positive
4. The overall consequences of killer robots are probably positive
5. Therefore, we should not ban killer robots

Given the caveats in 1, 2 and 4 above, I recommend the following policies (for details, see Müller and Simpson 2014):

6. Develop binding technical standards that spell out manufacturers' responsibilities
7. Maintain a clear chain of command and collect data, assuring responsibility for actions and enhancing provability of war crimes
8. Affirm and defend the humanitarian law 'just war' requirements, esp. clear differences between war and peace, between combatants and civilians (distinction) and between necessary and unnecessary force (proportionality). These requirements are under threat with current remote-controlled weapons, and this threat will continue with LAWS.
9. Do not ban lethal autonomous weapon systems (LAWS) or 'killer robots'

References

- Anderson, Kenneth and Waxman, Matthew (2013), 'Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can', *Stanford University, The Hoover Institution*.
<http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2250126%3E.
- Arkin, Ronald C (2009), *Governing lethal behavior in autonomous robots* (Boca Raton: CRC Press) pp.
- (2013), 'The Robot didn't do it: A position paper for the Workshop on Anticipatory Ethics, Responsibility and Artificial Agents',
<http://www.cc.gatech.edu/ai/robot-lab/online-publications/position_paper3.pdf%3E.
- Battaglia, Fiorella; Mukerji, Nikil and Nida-Rümelin, Julian (eds.) (2014), *Rethinking responsibility in science and technology* (RoboLaw Series, Pisa: Pisa University Press). 216
- Bertolini, Andrea (2014), 'Robots and liability - Justifying a change in perspective', in Fiorella Battaglia; Nikil Mukerji and Julian Nida-Rümelin (eds.), *Rethinking responsibility in science and technology* (RoboLaw Series; Pisa: Pisa University Press), 143-66.
- Birmingham Policy Commission (2014), 'The security impact of drones: Challenges and opportunities for the UK', (October 2014).
<<http://www.birmingham.ac.uk/research/impact/policy-commissions/remote-warfare/index.aspx>%3E.
- Emmerson, Ben (2014), 'Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism', *UN General Assembly, Human Rights Council*, 11 March 2014 (A/HRC/25/59).
<http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/25/59%3E.
- Enemark, Christian (2014), *Armed drones and the ethics of war: Military virtue in a post-heroic age* (War, Conflict and Ethics; London: Routledge) pp.
- European Parliament (2014), 'Plenary Session: Joint motion for a resolution on the use of armed drones', 25.02.2014 [adopted 27.02.2014] (2014/2567(RSP)).
<<http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P7-TA-2014-0172&language=EN>%3E.
- Garcia, Denise (2014), 'The case against killer robots: Why the United States should ban them', *Foreign Affairs*, (10.05.2014).
<<http://www.foreignaffairs.com/articles/141407/denise-garcia/the-case-against-killer-robots>%3E.
- Gunkel, David and Bryson, Joanna (eds.) (2014), *Special Issue on Machine Morality* (Philosophy & Technology, 27;
- Heyns, Christof (2013), 'Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions', *UN General Assembly, Human Rights Council*, 23 (3), A/HRC/23/47.
- ISS (2014), 'International Institute for Strategic Studies: The Military Balance 2014 (Press Statement)'.
<http://www.iiss.org/-/media/Documents/Publications/The_Military_Balance/MB14_Press_Statement.pdf%3E.

- Krishnan, Armin (2009), *Killer robots: Legality and ethicality of autonomous weapons* (Aldershot: Ashgate) pp.
- Leveringhaus, Alex and Giacca, Gilles (forthcoming 2014), 'Robo-Wars: The Regulation of Robotic Weapons', *Oxford Martin School Policy Papers*.
- Lichocki, Pawel; Kahn, Peter and Billard, Aude (2011), 'The Ethical Landscape in Robotics', *IEEE Robotics and Automation Magazine*, 18 (1), 39-50.
- Markoff, John (2010), 'War Machines: Recruiting Robots for Combat', *The New York Times*, (27.11.2010).
<<http://www.nytimes.com/2010/11/28/science/28robot.html>>
- Ministry of Defense, Chiefs of Staff (2011), 'The UK approach to unmanned aircraft systems', *Joint Doctrine Note*, 2/11 (30.03.2011).
- Müller, Vincent C. (2012), 'Autonomous cognitive systems in real-world environments: Less control, more flexibility and better interaction', *Cognitive Computation*, 4 (3), 212-15.
- (2015), 'Gun control: A European perspective', *Essays in Philosophy*, 16 (2), 247-61.
- Müller, Vincent C. and Simpson, Thomas W (2014), 'Killer robots: Regulate, don't ban', *University of Oxford, Blavatnik School of Government Policy Memo*, November 2014, 1-4.
- Müller, Vincent C. and Bostrom, Nick (2016), 'Future progress in artificial intelligence: A survey of expert opinion', in Vincent C. Müller (ed.), *Fundamental Issues of Artificial Intelligence* (Synthese Library; Berlin: Springer).
- Palmerini, Erica, et al. (2014), *Guidelines on Regulating Robotics*, 22/09/2014 (<http://www.robolaw.eu/>) 215 pp.
- Schmitt, Michael N. (2014), 'A Step in the Wrong Direction: The EU Parliament's Drone Resolution', *Just Security*, 07.03.2014.
<<http://justsecurity.org/7926/step-wrong-direction-eu-parliaments-drone-resolution-schmitt/>>
- Sharkey, Noel E (2008a), 'Computer Science: The ethical frontiers of robotics', *Science*, 322 (5909), 1800-01.
- (2008b), 'Grounds for discrimination: autonomous robot weapons', *RUSI Defence Systems*, 11 (2), 86–89.
- (2012), 'Automating Warfare: lessons learned from the drones', *Journal of Law, Information & Science*, 21 (2).
- Simon-Michel, Jean-Hugues (2014), 'Report of the 2014 informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)', *High Contracting Parties to the Geneva Convention at the United Nations*, (16.04.2014), 1-5.
- Simpson, Thomas W and Müller, Vincent C. (forthcoming), 'Just war and robots' killings', *The Philosophical Quarterly*, (online 07.08.2015).
- Singer, Peter (2009a), 'Military robots and the laws of war', *The New Atlantis*, Winter 2009. <<http://www.thenewatlantis.com/publications/military-robots-and-the-laws-of-war>>
- (2009b), *Wired for war: The robotics revolution and conflict in the 21st Century* (New York: Penguin) pp.
- Smalley, David (2014), 'The future is now: Navy's autonomous swarmboats can overwhelm adversaries', *US Office of Naval Research*, 05.10.2014.

<http://www.onr.navy.mil/en/Media-Center/Press-Releases/2014/autonomous-swarm-boat-unmanned-caracas.aspx>

United Nations, General Assembly (2013), 'Resolution: Protection of human rights and fundamental freedoms while countering terrorism', 68/178 (18 December 2013). http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/68/178

US Department of Defense (2012), 'Directive 3000.09, Autonomy in weapon systems'. [2014 Killer Robots Policy Paper Final.docx](#).

— (2013), 'Unmanned Systems Integrated Road Map FY2013-2038'. <http://www.defense.gov/pubs/DOD-USRM-2013.pdf>