The Ethical Troubles of Future Warfare. On the Prohibition of Autonomous Weapon Systems

Mihail-Valentin Cernea

ANNALS of the University of Bucharest
Philosophy Series

Vol. LXVI, no. 2, 2017
pp. 67–89.

THE ETHICAL TROUBLES OF FUTURE WARFARE. ON THE PROHIBITION OF AUTONOMOUS WEAPON SYSTEMS

MIHAIL-VALENTIN CERNEA[1]

Abstract

This paper is concerned with evaluating the arguments given to support the prohibition of autonomous weapon systems (AWS). I begin by offering a definition of autonomous weapons systems, focusing on the kind of autonomy involved by this type of combat robots. I continue by exploring Ronald Arkin's main arguments for ethical advantages in warfare that could be gained by the development and use of AWS (larger change of real world conflicts to actually comply with the international laws of war). The main part of the paper is dedicated to appraising what kinds of prohibition the international community can impose on such advanced weaponized robots and the kinds of arguments given by the proponents of such a ban. I propose a threefold classification of the arguments: epistemic, consequentialist and deontological. Of these three types of arguments, I argue that deontological arguments are the weakest, given the fact that their requirements are not satisfied by most weapons employed in war and that consequentialist arguments are more convincing if we are to ban the development of AWS. Regarding epistemic arguments and the legal arguments based upon them, they can be used to prohibit the use of AWS, but they seem to be neutral regarding the elaboration of these Artificial Intelligence based warfare technologies.

Keywords: drones, ethics of war, autonomous weapon systems, killer robots, just war theory, ICRAC.

1. Introduction

Popular culture is brimming with apocalyptic scenarios of a future in which Artificial Intelligence (A.I.) controlled robots murder human beings indiscriminately, for reason ranging from the mysterious to the

[1] "Alexandru Ioan Cuza" University of Iași. Email: cernea.mihai@gmail.com.

malevolent. The rise in popularity of what we call drones (unmanned combat aerial vehicles or UCAVs in military lingo) in America's war against terrorism coupled with the fascinating evolution of artificial intelligence has promoted the aforementioned scenarios from geeky fantasies about the future to a possible reality that is maybe just a few decades away from coming about. This is why a discussion about the morality of autonomous weapons systems is necessary – it will inform those that develop such systems about the failsafes and laws they should program in the A.I. controlled drones of the future, it will, we hope, shape the way state and non-state actors use of these weapons should a violent conflict begin and if they are permitted to do so.

Academic practical ethics is all over this subject and, unlike so many other moral issues discussed today, there seems to be a general consensus among military-minded ethicists around the world on this issue: we should ban their development, deployment and use (Sharkey 2011, 2012, 2017; Tamburrini 2016; Coeckelbergh 2016; Asaro 2012; ICRAC and other international organizations also support a ban or at least a moratorium). This paper is concerned with the kinds of arguments used when proposing this prohibition and their effectiveness in arguing for the ban envisaged by so many ethicists. I propose we classify the ethical arguments against autonomous weapon systems as follows: epistemic arguments (concerned with the commonly known problems of the development of artificial intelligence preclude the development of such autonomous killing robots), consequentialist arguments (concerned with showing why the outcomes of such technological developments are undesirable) and deontological arguments (concerned with showing why in principle we should never allow machines to select targets for the use of force because of human rights issues). These three types of arguments can be used to construct various legal arguments the promote a moratorium or an outright ban on autonomous weapons systems (these would be concerned with future machines' inability to comply with International Humanitarian Law, the international criminal law or any other legal system that governs conduct in war).

The article is structured as follows: firstly, I delineate what kind of military equipment is the target of the moratorium or outright ban;

secondly, I present some of the arguments offered for the development and use of autonomous weapons systems; thirdly, I critically survey the great diversity of arguments offered against the development and use of autonomous weapons systems under the broad classification that I have just sketched; lastly, I end the paper favoring epistemic arguments over deontological ones as grounds for a moratorium or ban on automated weapon systems.

2. Not All "Killer Robots" Are Created Equal. A Discussion of Autonomy

Before evaluating the ethical arguments surrounding autonomous weapons systems (I will refer to them through the acronym AWS for the remainder of the paper) we need to clearly understand what they are. This is not a paper about the moral challenges of contemporary drone warfare, but only about those future drones capable of selecting and eliminating enemy targets based on a preprogrammed ruleset, without the need of any human intervention.

There has been a lot of confusion around this subject mainly because popular culture, but also their designers have decided unwittingly to humanize these machines. Many press outlets call them killer-robots or killer-drones, and the names of UCAVs currently deployed by the United States in their active warzone also have been named in ways that are confusing to the general public. Examples include unfortunate metaphorical names like Reaper and Predator, names associated in most western countries with death, killer aliens and other sci-fi warfare technology (Bergman 2016, 178). While the use of these drones has interesting ethical aspects of its own, Reaper and Predator UCAVs are controlled by human beings at all times and do not offer the same ethical and technical challenges that future AWS will and, as such, are not in the scope of this paper.

First and foremost, it is important to grasp what it means to be autonomous in this context. An autonomous system is not the same thing as an automatic system.

> An automatic system is a fully preprogrammed system that can perform a preprogrammed assignment on its own. Automation also includes aspects like automatic flight stabilization. Autonomous systems, on the other hand, can deal with unexpected situations by using a preprogrammed ruleset to help them make choices. (Vergouw *et al.* 2016, 26)

The automatic system cannot choose to act in a certain way, it just performs the same task, repeatedly, indifferent of changes to its environment. An autonomous system, on the other hand, can adapt to its environment within certain limits set by its designer.

Moving on, it is useful to distinguish between levels of autonomy and types of autonomy. When we talk about levels of autonomy, we refer to the level of disconnect the drone has from its human operator, if it has any. To take a real world example, the United States Department of Defense (DoD) outlines four levels of autonomy in the a document detailing the future development of land, air and sea based drones: (i) no autonomy – all decision making regarding drone operation is made by the human operator; (ii) delegated autonomy – the human operator directs the system to perform certain tasks, but no further input is required (automation); (iii) supervised autonomy – the autonomous system can initiate task based on sensory input, not only the human operator; (iv) full autonomy – the human operator only inputs general commands, while the system selects and performs the tasks related to the commands. The DoD necessitates that the human operator can always intervene in all types of autonomy, in case of an emergency (USDoD 2013, 26-28; Vergouw *et al*. 2016, 25-26).

When we talk about types of autonomy, we refer to a distinction between personal autonomy and task autonomy. Personal autonomy is related to conscious individuals capable of pursuing their self-imposed interests, while task autonomy is related to a certain system capability to perform a certain task without any outside help (Tamburrini 2016, 125). As it is quite obvious from the current status of A.I. research, it is quite unlikely that any realistic projection about AWS in the near future involves personal autonomy. More useful for our contemporary case is the idea of task autonomy. The performance of the task is essential here, not the demonstration of having a self. An AWS does not need be a complex artificial sentient being as those shown in various sci-fi books

and movies, but only capable of performing its task independent of human intervention. Furthermore, task autonomy does not entail total autonomy. For example, a cat may be task-autonomous from humans in finding and hunting its food, but may well need the help of a fireman to get down from a tree. An AWS does not depend on humans to find its way around the battlefield, but may depend on other systems, like GPS or GLONASS for this task. Again, the US Department of Defense provides us with a good working definition for what an AWS is, from the point of view of this distinction: a weapon system is autonomous if, "once activated, can select and engage targets without further intervention by a human operator" (USDoD 2012, 13-14). As one can clearly see in this case, the AWS envisaged by military forces of the United States is defined around task-autonomy, not personal autonomy. One does not need to be HAL 3000 to be an AWS. What is essential for the ethical discussion around AWS will be related, in my own view, to the idea of machines capable of selecting and eliminating enemy combatants without *human* intervention. Whether this is done by an advanced A.I. like HAL 3000 (of Kubrick's *2001: A Space Odyssey*) whose inner workings are mysterious even for our imagination or just a simple ruleset that identifies a certain uniform (in the case of conventional war) or statistical learning based on enemy behavior (as I would imagine would happen in less conventional battlefields, like those of Afghanistan) is not very important at this stage – most authors emphasize humans being out of the loop, as we shall see in the fourth section of this paper.

Guglielmo Tamburrini adds another element to the concept of task-autonomy. Some AWS may work in clear cut environments, but not in messier combat theaters. Tamburrini's example is the Korean robotic sentinel SGR-A1 that can autonomously survey and fire upon targets in the Korean Demilitarized Zone (DMZ), but would struggle to make correct decision without human intervention in an urban warfare scenario. "This observation suggests that t-autonomy should be more accurately construed as a relationship between four elements: a system S, a task t, a system S′ from which S does not depend to accomplish t and an environment where t must be performed" (Tamburrini 2016, 127). This adds another layer of problems, ethical and technical, for developers and

operators of AWS, as it is unclear whether they will work up to standard on the unfortunately many and diverse battlefields of planet Earth. In Tamburrini's example, the Korean DMZ is easy to process for the SGR-A1, as humans are strictly prohibited from entering the neutral area between the divided Koreas. Target acquisition is not precluded by the presence of civilians and other disruptive elements that would otherwise be present in a theatre of operations like Fallujah, Raqqa or Aleppo. These kinds of obstacles will feature prominently in the epistemic arguments for a moratorium on AWS that I will examine in the fourth section of this paper.

To conclude, the autonomous weapon systems are robotic platforms capable of selecting and engaging enemy combatants and other targets of military value without the direct intervention of human operators. While human operators may be able to override the machines' decisions, it is important to underline the fact that for a drone or a similar type of weaponized robot to be an AWS it needs to be autonomous from human agents regarding the task of using force. Another important fact to note is that AWS don't need to be aerial vehicles, even though contemporary discussion is concentrated on UCAVs. AWS can be air-based, sea-based or land-based. Moreover, they can be static defense systems – in this case the ethical difficulties are lessened by the fact that static defense systems do not seek and destroy and are usually programmed to protect friendly territory from enemy ballistic missiles. In the rest of the paper, my concern will be with mobile AWS.

3. Robots may not Feel Empathy, but Robots Don't Feel Hatred as Well. The Main Argument for AWS

Before taking a look at arguments against the development and use of AWS, it is useful to understand why some military experts and even some ethicists have manifested a certain enthusiasm for a future where complex ethical decision-making in war is done by robots rather than soldiers, pilots or even commanders.

3.1. Just War Theory and AWS

The ethical framework governing military conflicts is just war theory and the debate surveyed in this paper will be seen through this theoretical lens. Simply put, as a deep discussion of whether there are such things as just wars is beyond the scope of this article, just war theory is concerned with the condition in which the use of force by state or non-state actors is morally permissible, but not necessarily justified or required. For the remainder of the paper I will assume that there are such things as just wars and, despite debates between authors in the just war tradition, this framework works in giving us an ethics of armed conflict. There are three set of principles that just war theory provides for the use of military force:

(i) *Jus ad bellum* – this set governs the resort to war. To resort to war a state or non-state actor is usually required to fulfill the following criteria: having a just cause, using a proportional response, the use of force being necessary and last resort, having legitimacy and reasonable chance to win (Leveringhaus 2016,12; McMahan 2009, 3-4).

(ii) *Jus in bello* – this set governs the conduct of armed forces during war. The main guidelines for ethical behavior in combat are concerned with the distinction between combatants and non-combatants, the weapons involved (don't bring a nuke to a fist-fight!) and whether the battle is actually necessary. (Leveringhaus 2016,12; McMahan 2009, 4-5).

(iii) *Jus post bellum* – this set governs behavior in the aftermath of the war and is a more recent addition to just war theory that is still hotly debated (Leveringhaus 2016, 13; Orend 2000).

Intuitively, we can see that the three sets of principles are rather independent from each other: one can fight a just war with unjust means or the other way around, one can win a just war through just means and still have an unethical behavior in the aftermath of the war and so on and so forth.

Where do AWS fit in this large ethical framework that I have sketched here? The literature I have reviewed for this paper stands in a general agreement that the main ethical challenges for the development

and deployment on the battlefield of autonomous weapons systems come from the *jus in bello* set of principles. Some authors also believe that the use of AWS also merits a discussion on its compliance to *jus ad bellum* principles. Heather Roff, for example, argues that deploying AWS even in the most just of wars would affect proportionality and would also lead to a new arms race (Roff 2015). The main issues that supporters and critics of AWS alike seem to underline (no matter the kinds of arguments preferred) regard the capabilities of autonomous robotic soldiers to distinguish between combatants and non-combatants in real world warfare scenarios and the weapons to be employed. It is not the robots that decide to go to war (at least for the foreseeable future) and it is not the military robotic platforms that must act justly in the aftermath of a war (though civilian robotic applications would probably play a role).

3.2. Ronald Arkin's Arguments for the Development of AWS

The main proponent of AWS is Ronald Arkin. It may seem counterintuitive at first, but the fact that AWS do not have feelings, do not tire and do not 'understand' the harm they may inflict on their human targets is a feature of AWS rather than a caveat. The gist of the argument is that AWS would be more likely to respect the principles of *jus in bello* and the International Humanitarian Laws the govern military conflict more readily than human as they lack the traits that predispose humans to unethical behavior in war. Anger and fear of violence, grief because of the loss of loved friends and other psychological and physical stress factors of combat often drive even the most resilient soldiers to commit heinous war crimes (Arkin 2009). Some examples come to mind: many historians of World War I underline how the conflict became more and more brutal as time went by, as soldiers lost comrades, as civilians suffered at the hands of the enemy, as more brutal weapons were deployed and as living conditions of the Great War's many fronts degraded. Machines programmed to respect the International Humanitarian Law have no option but to do so and will not seek revenge on innocent civilians or surrendering enemy combatants.

Moreover, Arkin also argues that machines would do better in wars not only because they are not subject to the darker demons of our nature, but also because they would, ideally, lack the epistemic limitations that humans must contend with: AWS would not suffer from the cognitive biases that affect us nor from the hardwired information processing limitations that human beings have. It's also worth noting an interesting point made by David Bergman about arguments similar to those of Ronald Arkin that are being made today for driverless cars:

> It is humans that make mistakes and violate traffic rules or exceed the speed limit. Thus we should rationally trust that an autonomous car would follow traffic rules better than a human. This leads to an important question. If we can trust a fully autonomous car to follow the traffic rules better than a human, should we not also be able to trust a fully autonomous drone to perform better than humans in complying with ethical and legal codes of conduct on the battlefield as well? (Bergman 2016, 177).

A.I. drivers would not try cut in line, would not be tempted to cross a red light and would not be involved in road-rage incidents. Same principles may work for A.I. soldiers, as long as it is technically feasible.

Arkin's claims are echoed by officials from the US military who claim that their A.I. combatants "don't get hungry, they're not afraid. They don't forget their orders. They don't care if the guy next to them has just been shot. Will they do a better job than humans? Yes." – Gordon Johnson of the Joint Forces Command at the Pentagon in 2005, earlier than Arkin's influential paper on this subject (Weiner 2005).

4. No Robots Allowed Past This Ethical Point. Ways to Argue Against AWS

Many philosophers do not share Arkin's enthusiasm for AWS, as will be shown by the remainder of this paper. Moving in the opposite direction, in recent years, much of the scholarly literature I have taken into account has called for a moratorium or ban the development and/or deployment of such weapons. Images of swarms formed out of autonomous amoral killer drones storming battlefields and eliminating

targets left and right without any oversight have frightened the imagination of laypeople and ethicist alike, but the reasons invoked by AWS's enemies are many and pose real ethical challenges outside the realm of emotion and imagination. This is not just technophobic scaremongering as some might argue, but real concern regarding a piece of technology that could make literal life and death decisions in the most difficult landscape of human experience, war.

4.1. What to Ban?

First of all, we need to distinguish the three ways in which one could put in effect an AWS ban. One could ban the future development of AWS – for this ban to be justified, one needs to show that there are no possible conditions under which the mere existence of such advanced robotic weaponry would be morally justified. To further complicate matters, one could ban any research that could lead to AWS, which is unrealistic, given that portions of the technology require for autonomous weapons systems are already a part of civilian life, or one could ban only the research required to build the actual 'killer' robot. For such a ban to make sense, either deontological arguments need to be given which show why such weaponized artificial beings would violate certain generally accepted ethical principles or a more consequentialist approach that warns against the negative effects of such research. Another option available to adversaries of AWS is to ban the production of said machines. In this case, my opinion is that arguments need to strike at the consequences of nation states and non-state actors of AWS. Like in the very unfortunate case of chemical weapons, the research may be out there, but firm red lines regarding the possession AWS may be enforced. A similar case could be built for a ban on the deployment of AWS on the battlefield. Here too one could see a nuance based on Tamburrini's discussion of the variable accuracy of AWS depending on the environment in which they are deployed: rather than just banning the use of AWS outright, international treaties could just prohibit autonomous killing machines in the environments where their effectiveness has not been demonstrated beyond reasonable doubt.

Any proposal for a moratorium or ban on AWS should also take into account, from my point of view, what type of prohibition would we realistically could expect to work. Any international treaty would affect only those actors that are submitted to it. A terrorist organization that has the know-how necessary to develop autonomous military robots (as technology develops, A.I. platforms get better and cheaper, such a possibility may not be as far-fetched as it seems now) may find itself at a copious advantage in the face of enemy nation states that have successfully restrained themselves from even developing the tech for AWS.

4.2. Types of Arguments against AWS

In this subsection of the paper, I propose a useful working classification of the arguments given for a moratorium or ban on AWS. As it is to be expected, not every argument may fall neatly in the three types I have tried to delineate in the following few pages. This being said, I think there roughly three ways in which most authors argue against AWS:

(i) Epistemic arguments – these are arguments that purport to show that we do not know nearly enough about the various technological elements of AWS or the way they would function in a real life armed conflict so as to approve their use. Not much ethics here, but there are ethical conclusions to be drawn.

(ii) Consequentialist arguments – these are arguments that assume a consequentialist meta-ethical framework. The main idea in these arguments goes along the following lines: here are negative consequences of developing/producing/using AWS, thus it is immoral to develop/produce/use AWS.

(iii) Deontological arguments – these are arguments that maintain some aspect of autonomous drone warfare would breach one or more ethical principles assumed by the internationally recognized framework of human rights. Usually, the basic idea is that allowing a machine to decide whether a human being must live or die is a moral wrong in and of itself.

Based on the three kinds delineated above one can build legal arguments for a ban or moratorium on AWS, as they all show they cannot comply with various aspects of the international law on conflicts. While strong, as we are about to see, the validity of these arguments depend on the validity of the philosophical arguments they stand on. For example, as long as the technology "is not there yet", there are no guarantees that AWS will respect the International Humanitarian Law

Now, I will critically examine each category in part in order to better understand their role in future of AWS development.

4.2.1. Epistemic Arguments

This class of arguments against the A.I. controlled warfare robots basically shows us that we are far away from being able to employ mobile AWS in battle. There are two main reasons from what I can gather: one is related to the epistemic difficulties posed by 'unstructured warfare scenarios', while the other is related to the fact that current A.I. research has serious issues in understanding exactly how do the statistical learning algorithms behind most contemporary make the decisions they do.

Tamburrini argues that contemporary civilian robotic applications depend for the achievement of their tasks on "adapting environment to the perceptual, cognitive and action capabilities of robotic systems" (Tamburrini 2016, 128). One example are industrial assembly lines in which automated robots are segregated from their unpredictable human colleagues so as not to perturb their orderly function. Have one variable of that environment change in a significant way and tremendous errors may occur.

Even in the case of robots that do have interact with humans, like personal care robots that are meant to assist the elderly and the disabled, strict guidelines, ranging from constraining the operational scenarios to warning labels meant to educate the users on the dangers of unstructured interaction, need to be followed so that accidents do not occur. As Robert Sparrow argues, "despite many decades of research – and much progress in recent years – perception remains one of the *hard*

problems of engineering. It is notoriously difficult for a computer to reliably identify objects of interest within a given environment and to distinguish different classes of objects" (Sparrow 2016, 8).

This hard problem of machine perception is hundredfold worse in the unstructured environments of contemporary war. Especially, one might add, in the modern war on terror where the lines between combatant and non-combatant are infinitely blurred. Let's imagine AWS in Afghanistan where "not every person carrying a weapon is a combatant" (Sparrow 2016, 9) and not every person that does not carry a weapon is a civilian. When A.I. controlled robots face such great hurdles in quiet elderly homes, let us just think about how much damage they could do to innocents in contemporary warzones.

The other issue I mentioned is connected to the statistical learning models underlying most A.I. technology used in contemporary applications ranging from driverless cars to language translation. This is the kind of artificial intelligence that would be used in target selection for future AWS. The problem with it: its creators do not understand exactly what is behind its decision making process because of the complexity of the neural networks involved. As this kind of machine learning is built based on human neural networks, which to this day pose difficult problems to neuroscientists and philosophers of the mind, it is easy to see why "there is no obvious way to design such a system so that it could always explain why it did what it did" (Knight 2017). If we don't understand (yet!) how the human mind works, we will have similar difficulties in understanding artificial minds built on similar 'hardware'[2]. In warfare, this raises big questions of accountability and correction. When contemporary drone strikes go wrong (like when they accidentally blow up a wedding procession thinking it is a terrorist cell), it is, more often than not, a case of bad intelligence: either a source lied, an intelligence officer did not do her job or other causes that are easy to understand. If an AWS destroys a wedding party, it may be impossibly

[2] I do not mean to attach myself to any philosophical perspective on the mind with this comment. Rather, I would like to underline that A.I. research may find itself against epistemic obstacles similar to those that have precluded philosophical and scientific investigations of consciousness in the past and present.

difficult to grasp why did go so wrong so that we can avoid such horrendous situations in the future.

The epistemic arguments show that AWS development is actually quite far off. As long as these technical hurdles remain, there is good reason to prohibit the production and deployment of AWS.

4.2.2. Consequentialist Arguments

Arkin's arguments for the use of AWS are of a consequentialist nature: they maintain that the deployment of AWS would lessen the risk of friendly troops, increase precision in targeting legitimate targets and ensure compliance with International Humanitarian Law, thus it would be ethical to use them. As we have seen in the previous section, there are significant technological obstacles in achieving the said positive consequences.

Scholars like Jürgen Altmann and Tamburrini propose a distinction between narrow consequentialist arguments that do support the development and use of AWS and wide consequentialist arguments that underline the future problems of the proliferation of AWS – "one would significantly raise the risk of a new arms race and global destabilization, by providing incentives for the commencement of wars and by weakening traditional nuclear deterrence factors based on mutually assured destruction" (Tamburrini 2016, 138). Using AWS would decrease the political risk associated with starting armed conflicts for world leaders. Simply put, when the troops are not at risk and one has available a large supply of robots that have no families of voters, then the incentives to start wars rise. AWS operations can take higher risks, going as far as even to dismantle a nuclear power's strike capabilities, thus stimulating risky first strike scenarios and bringing unpredictability to our world. In this situation, investment in AWS would be recommended for all state actors, in an attempt to regain the equilibrium of M.A.D. This means increased risk of AWS warfare, but also increased national military budgets which could preclude a state's capacity to invest its tax revenue in other worthy social causes.

Another serious risk we can associate with the development and use of AWS is the risk of hacking. Given capable enough hackers, fleets

of AWS can be directed to attack friendly forces in war and terrorize civilians in times of peace (Leveringhaus 2016, 120-121).

The epistemic arguments quoted above show that it may be too early to know the effects of AWS, but this strict consequentialist discussion indicates that, in the end, the bad consequences may far outweigh the positive consequences of automated warfare.

4.2.3. Deontological Arguments

When it comes to killing humans in war or otherwise, irrespective of whatever ghastly means we may choose to complete this grievous task, other humans need to be 'in the loop' if there's to be even the faintest chance that our actions be morally permissible. Autonomous weapon systems seem to lack the capacity to achieve this moral goal, or so the argument maintains. The basic idea is that AWS infringe the basic principles of human rights. A human may only be killed if another human has carefully considered that person's case through due process and not by an automated process that is unaccountable even if, at least prima facie, if respects the International Humanitarian Law.

This type of argument may try to achieve its normative goals in two related, but different ways: either one shows that IHL and just war theory require that a human be 'in the loop' whenever there's a decision to use lethal force (Asaro 2012, 688-689; Horowitz 2016; Sparrow 2007) or by questioning the moral nature of machines and demonstrating that, in principle, they lack the moral capacity (they cannot be moral agents, but, more interesting, they cannot be moral patients, so they can't understand the damage they are inflicting on their human targets) to decide whether a person should live or die, regardless of their technological sophistication (Coeckelbergh 2016, 234-235). Alex Leveringhaus' Argument from Human Agency is a special case as it combines the two deontological approaches mentioned: "Human agency, I argue, entails the ability to do otherwise by engaging in an alternative course of action. In a nutshell, soldiers have the ability not to pull the trigger, while a machine that been pre-programmed does not. Unless re-programmed, the machine will engage the targeted person

upon detection. Killing a person, however, is a truly existential choice that each soldier needs to justify before his own conscience." (Leveringhaus 2016, 92) Humans need to be in the loop, because machines lack the basic ability to refuse an immoral order or to have mercy (sometimes enemy combatants need only be disabled, not outright killed).

While deontological arguments that focus on the lacking moral capacities of AWS could make the strongest case for prohibiting even the development of AWS, in my view they assume an ideal world that is incompatible with the unfortunate ways in which our actual world works. There are few current practices of war that, even with humans involved in the decision to use lethal force, can satisfy them. Deontological arguments make most kinds of killing in war immoral. Consider the case of bombing raids in the two World Wars that shook up the 20th Century: yes, humans are piloting the bombers and dropping the payloads, but in the actual killing of human targets what decides who lives or who dies is also determined by chance: war is not an exact science, as working with explosions will always carry a certain factor of randomness that will affect who lives or who dies regardless of how many humans are in the loop. Today's battlefields have not overcome this problem, even if carpet bombing is now prohibited by the Geneva Convention: grenades, rocket-propelled grenades, artillery and airstrikes cannot escape this factor of imprecision that moves the whole enterprise of killing human in war outside margins that humans can reliably control in the ways required by these types of arguments. AWS are just another case of humans using imprecise and means to kill other humans, even they are imprecise for different reasons than an RPG fired in the general direction of the enemy.

Moreover, what I have stated above may show that humans are in the loop in the same manner they are in the loop in contemporary conventional war. A commander decides whether to deploy AWS on the battlefield, knowing full well the risks of using autonomous decision processes to kill other humans. Thus, the responsibility for any unjustified deaths is that of the commander, because she carries the moral traits required by the deontological arguments stated above. A similar point has been made recently by Champagne and Tonkens

regarding the "accountability gap" in autonomous warfare. They propose to bridge this gap through what they call "blank check" responsibility – occupying the office that governs the use of AWS could mean assuming the responsibility of their actions indifferent of any causal relation to the immoral pains that could be inflicted with their deployment during the battle. "In essence, our proposal retains the non-causal imputation involved in scapegoating while dropping its arbitrariness: since humans are capable of informed consent and pleasure/pain, a suitable and ascertainable target for punishment can be established, thereby ensuring visible conformity with the tenets of just war theory. While victims of the immoral behavior of autonomous lethal robots may not always be satisfied with the level of retribution gained (*e.g.*, the forfeiture of office/rank, fines, imprisonment), it is important that punishment for such actions go only to those that deserve it, and getting some fair retribution is better than not getting any at all" (Champagne and Tonkens 2015, 136).

Taken to their conclusion, deontological arguments like those stated above show us why war is in general immoral and irresponsible, not just the use of AWS during battles. Countless violent historical events show us that the practice of war is an emergent state of affairs that almost always escapes our control (The Great War should have been over 'by Christmas', the useless bombing of Dresden and other many horrendous moments that haunt our shared history as human beings), bringing about innumerable innocent deaths. This has not stopped humans from starting wars, so there is a pragmatic need for things like the International Humanitarian Law or just war theory to limit the destruction we humans have a habit of inflicting upon each other even if using violence to achieve our goals is shown to be strictly immoral and shouldn't be encouraged or condoned by any body of laws. In the same way, if AWS achieve the pragmatic goal of limiting human suffering in the stupid wars that we can't help ourselves to always start, irrespective of how immoral their deployment may be in the light of deontological principles, we should use them. Or stop going to war altogether.

This is why I believe that deontological arguments are not fit to justify any kind of legal action against autonomous killer robots. The

same grounds may justify the ban of any explosives and large scale operations. Maybe snipers who take their time in choosing their victims may escape the real scope of these arguments. If we are to ban the development and use of AWS, we need grounds of a different nature. Simple normativity is, at the same time, too much and too little.

4.2.4. Legal Arguments

The main problem that is outlined by these arguments is the expected failure of AWS to comply with the International Humanitarian Law, that is, the law of armed conflict. The basic idea is that autonomous weapons would not be able align with the principles of distinction, proportionality and precaution. I believe we can show that these kinds of arguments are valid only by appealing on epistemic and consequentialist grounds, escaping the more difficult philosophical (and ontological) discussions engendered by normative grounds.

Regarding the principle of distinction, the epistemic arguments outlined in section 4.2.1 show the machine perception is not able to reliably distinguish between civilian and belligerent targets in the ever shifting environments of war. "Another problem for the Principle of Distinction is that we do not have an adequate definition of civilian that we can translate into computer code" (Sharkey 2017, 179). The laws are quite vague on what a civilian is: in 1949 the Geneva Convention recommended the use of common sense, while the 1977 Additional Protocol I essentially described a civilian as a non-combatant. It is quite hard to see what algorithms could be built in order to ensure machine would know, even with perfect machine perception, that a certain individual on the battlefield is actually a civilian. Does that person hold a weapon? Well, maybe it's just a civilian robbing a dead soldier – robbing the dead may be immoral, but it is not the kind of offense that warrants the death penalty. It is unclear whether a machine could make that distinction. This problem becomes even more difficult when a 'killer robot' is supposed to distinguish between civilians and civilians directly participating in hostilities, which can be attacked.

The principle of proportionality also comes with its own problems. Proportionality is not just about using the morally appropriate kind of weapon, but to balance the expected collateral casualties to the military advantage gained by the use of kinetic force. Also, given today's wartime challenges, the battle is not only fought to defeat the enemy, but to win the hearts and minds of the civilian populace so as to ensure the smallest possible likelihood of future armed conflicts in the territory affected by war. This involves cultural factors that are outside the capacity of contemporary robots (Sharkey 2017, 180).

Compliance with the principle of precaution suffers as well because, as of this moment, we are unable, because of the same epistemic reasons that affect the principle of distinction, to guarantee the predictability of AWS. "There are currently no formal methods available to determine the behavior of an autonomous system. It is still very difficult to formally verify anything but fairly simple programs and autonomous systems are still beyond the abilities of computer science; add learning algorithms and it gets orders of magnitude more difficult" (Sharkey 2017, 180).

To conclude, whether or not some law is respected is, in many ways, a practical epistemic issue. We must understand the real world requirements of the law (the objects and kinds of behavior that are underscored by that law) and the 'spirit' of the law (how we interpret that law and the reason it exists). For this reason, we must understand how AWS work and how their deployment will affect the human activity that is regulated by the IHL. Normative considerations are important, maybe most important in the philosophical grounding of the law, but the question the legal arguments in this context try to answer is whether AWS can comply with the law of armed conflict, not whether AWS have an ontological and ethical status that allows them to be subjects of these laws.

4.3. To Ban or to Temporarily Ban? That Is the Question

In September 2009, some of the authors quoted above (Altmann, Sharkey, Sparrow and Asaro) formed the International Committee for

Robot Arms Control (ICRAC) which proposes an international prohibition on autonomous weapons systems (Asaro 2012, 688). The ban involves the development, deployment and use of armed autonomous unmanned systems. Do their arguments support such large scale moratorium on AWS?

I think this question has actually two different components for which the argument classification I have proposed in the last subsections is relevant: the ban on development of AWS and the ban on use of AWS. Moreover, one needs to show whether the ban should be temporary (a moratorium) or should be a permanent ban.

The prohibition on the development of AWS should be grounded, in my view, on epistemic and consequentialist arguments. As I have stated above, I hold that deontological arguments suffer in the light of real life warfare. They state ideal conditions that cannot be satisfied by most technology used in armed conflict and if these is a chance that AWS could lessen the evil done by modern war, an outright ban on deontological grounds for their research and development would actually do more harm than good. If there is a provable way in which this kind of technological advancement might limit human suffering on the battlefield, given our planet's many wars, consequences should overtake in importance any deontological considerations. Wherever the discussion on the moral status of machines may take us, there is widespread agreement over the moral status of humans. If we need to permit humans not being completely in the loop when lethal force is used in war to lessen human casualties or we need to allow beings of dubious moral status to be involved in armed conflict to achieve the same goal, that is a philosophical cost that we should incur. The ideal case in which we know for sure that the deployment of AWS brings less violence and bloodshed in our battlefields is a situation in which various moral duties come into conflict: the duty regarding how morally permissible violence should only be decided by humans versus our duty to limit human suffering as much as possible. My view is that the latter is the more important one in this ideal case. If this is correct, then consequentialist and epistemic grounds are of primary importance when dealing with a possible moratorium or ban on AWS.

Given the frequent occurrence of armed conflict in humanity's history and the large scale death and destruction that comes with such events, a consequentialist meta-ethical framework is, in my view, much more useful in dealing with the moral quandaries of war. This is the reason that I believe the wide-consequentialist arguments offered by Altmann and Tamburrini are much stronger grounds for the ban on the development of AWS.

Regarding the prohibition of using AWS in battle, the epistemically grounded legal arguments are much stronger than the others, but only if one assumes the same consequentialist framework that I favor for the ethics of armed conflict. Paradoxically, if only the epistemic and legal arguments would affect the international community's stance on AWS, they would also work as arguments for the further development of AWS so that compliance with the laws of war be achieved. Even so, the consequences of a new and unpredictable robot arms race outweigh, in my mind, any benefits of using the ideal precise AWS that philosophers like Ronald Arkin envisage for the far future. Until we fulfil the epistemic requirements for the successful deployment of AWS, at least a moratorium could be justified.

5. Conclusion

The main purpose of this paper was to investigate whether a moratorium on the development and use of autonomous weapon systems is justified. To achieve this goal, I set out at first to understand what kinds of machines would be affected by such a ban. After arriving at a satisfactory meaning for the term "autonomy" envisioned by scholarly literature on this subject, I also mentioned that AWS need not be modeled after the UCAVs employed by the US military in its contemporary battlefields: AWS can be land, sea or air based.

The second part of the paper examined the main moral argument that grounds the future development and deployment of AWS: machines lack the darker side of human nature and they would be more likely to uphold the International Humanitarian Law and the ethical

guidelines of just war theory for the *jus in bello* set of principles that govern an army's conduct during combat operations.

The third and main part of the paper was dedicated to appraising what kinds of prohibition the international community can impose on such advanced weaponized robots and the kinds of arguments given by the proponents of such a ban. I proposed a threefold classification of the arguments: epistemic, consequentialist and deontological. Of these three types of arguments, I maintained that deontological arguments are the weakest, given the harsh realities of ethical choice during wartime and the consequentialist arguments are more convincing if we are to ban the development of AWS. Epistemic and legal arguments can be used to prohibit the use of AWS, but they seem to be neutral regarding the elaboration of these A.I. based warfare technologies.

REFERENCES

Arkin, Ronald (2009). "Ethical Robots in Warfare." *IEEE Technology and Society Magazine*, 9(4): 332-341.

Arkin, Ronald (2009). *Governing Lethal Behavior in Autonomous Robots*. Florida: CRC Press

Asaro, Peter (2012). "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making." *International Review of the Red Cross*, 94: 687-709.

Bergman, David (2016). "The Humanization of Drones: Psychological Implications on the Use of Lethal Autonomous Weapons Systems." In *The Future of Drone Use*, edited by Bart Custers, 173-188. The Hague: Springer.

Champagne, Marc and Tonkens, Ryan (2015). "Bridging the Responsibility Gap in Automated Warfare." *Philosophy & Technology*, 28(1): 125-137

Coeckelbergh, Mark (2016). "Drones, Morality, and Vulnerability: Two Arguments Against Automated Killing." In *The Future of Drone Use*, edited by Bart Custers, 229-239. The Hague: Springer.

Horowitz, Michael C. (2016). "The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons." *Daedalus*, 145(4): 25-36

Knight, Will (2017). "The Dark Secret at the Heart of AI." *MIT Technology Review*, April 11. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (accessed July 13, 2017).

Leveringhaus, Alex (2016). *Ethics and Autonomous Weapons.* London: Palgrave Macmillan.

McMahan, Jeff (2009). *Killing in War*. Oxford: Oxford University Press.

Roff, Heather (2015). "Lethal Autonomous Weapons and Jus Ad Bellum Proportionality." *Case Western Reserve Journal of International Law*, 47(1): 37-52.

Sharkey, Noel (2011). "The Automation and Proliferation of Military Drones and the Protection of Civilians." *Law, Innovation and Technology*, 3(2): 229-240.

Sharkey, Noel (2012). "The Evitability of Autonomous Robot Warfare." *International Review of the Red Cross*, 94: 787-799.

Sharkey, Noel (2017). "Why Robots Should not be Delegated with the Decision to Kill." *Connection Science*, 29(2): 177-186.

Sparrow, Robert (2016). "Robots and Respect: Assessing the Case against Autonomous Weapon Systems." *Ethics and International Affairs*, 30(1): 93-116.

Tamburrini, Guglielmo (2016). "On Banning Autonomous Weapons Systems: From Deontological to Wide Consequentialist Reasons." In *Autonomous Weapons Systems: Law, Ethics, Policy* edited by Nehal Bhuta, Susanne Beck, Robin Geiss, Hin-Yan Liu, Claus Kress, 122-142. Cambridge: Cambridge University Press.

United States Department of Defense (2012). *Directive 300.09: Autonomy in Weapons Systems*. http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf (accessed July 13, 2017).

United States Department of Defense (2013). *Unmanned Systems Integrated Roadmap*. https://www.defense.gov/Portals/1/Documents/pubs/DOD-USRM-2013.pdf (accessed July 13, 2017)

Vergouw, Bas and Nagel, Huub and Bondt Geert and Custers Bart (2016). "Drone Technology: Types, Payloads, Applications, Frequency Spectrum Issues and Future Developments." In *The Future of Drone Use*, edited by Bart Custers, 21-46. The Hague: Springer.

Wiener, Tim (2005). "New Model Army Soldier Rolls Closer to Battle." *The New York Times*. February 16. http://www.nytimes.com/2005/02/16/technology/new-model-army-soldierrolls-closer-to-battle.html (accessed July 13, 2017).