The Unfounded Bias Against Autonomous Weapons Systems

Autonomous Weapons Systems (AWS) have not gained a good reputation in the past. This attitude is odd if we look at the discussion of other – usually highly anticipated – AI-technologies, like autonomous vehicles (AVs); whereby even though these machines evoke very similar ethical issues, philosophers' attitudes towards them are constructive. In this article, I try to prove that there is an unjust bias against AWS because almost every argument against them is effective against AVs too. I start with the definition of "AWS." Then, I arrange my arguments by the Just War Theory (JWT), covering *jus ad bellum*, *jus in bello* and *jus post bellum* problems. Meanwhile, I draw attention to similar problems against other AI-technologies outside the JWT framework. Finally, I address an exception, as addressed by Duncan Purves, Ryan Jenkins and Bradley Strawser, who realized the unjustified double standard, and deliberately tried to construct a special argument which rules out only AWS.

Keywords: *artificial intelligence; autonomous weapons systems; self-driving cars; just war theory; AI ethics; military ethics*

Acknowledgements

This research was supported by the MTA Lendület Values and Science Research Group; and the UNKP-20-3 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

Author Information

Aron Dombrovszki, Eötvös Loránd University (ELTE); MTA Lendület Values and Science Research Group
https://elte.academia.edu/AronDombrovszki

How to cite this article:

Dombrovszki, Aron. "The Unfounded Bias Against Autonomous Weapons Systems." *Információs Társadalom* XXI, no. 2 (2021): 13–28.
https://dx.doi.org/10.22503/inftars.XXI.2021.2.2

Introduction

Autonomous Weapons Systems (AWS) have not gained a good reputation in the past: in 2012, Human Rights Watch suggested a pre-emptive ban, and The Campaign to Stop Killer Robots also raised concerns in the media with an aim to influence public opinion against AWS. The press usually refers to AWS pejoratively as "killer robots," accompanied by a picture of a frightening Terminator-like sci-fi machine (e.g. Kahn 2020; Kessel 2019; Scharre 2020). The academic debate on the topic, however, tends to be more balanced than the general discussion; nevertheless, the majority approach towards AWS is still negative (Rosert and Sauer 2018).

This prejudice differs markedly from the general attitude towards other artificial intelligence (AI) technologies, for example, autonomous vehicles (AVs). AVs are highly anticipated, and even though similar worries can be raised against them, philosophers' attitude is far more accepting.

Purves, Jenkins and Strawser (2015) also mention this kind of a double standard between AWS and other AI-technologies:

> Any account of the permissibility of autonomous weapons systems will risk prohibiting the use of autonomous decision-making technologies that most people view as neutral or morally good. While many of us tend to have a significant moral aversion to the thought of autonomous weapon systems, most have no such similar moral aversion to non-weaponized autonomous systems, such as driverless cars. In fact, for many people, the opposite is true: many of us hold that non-weaponized future autonomous technology holds the potential for great good in the world.

In this article, I aim to prove that Purves, Jenkins and Straswer's observation is accurate, and almost every argument against AWS is effective against AV. To achieve this goal, I start with the definition of "AWS". I then arrange my arguments by the Just War Theory (JWT), covering *jus ad bellum, jus in bello* and *jus post bellum* problems against AWS. Meanwhile, I draw attention to similar problems against other AI-technologies outside the JWT framework. The aforementioned Purves, Jenkins and Strawser article is an interesting exception, because they realized the unjustified double standard, and deliberately tried to construct a special argument which rules out only AWS. I conclude by addressing their arguments and show their vulnerabilities.

1. What are AWS?

Before discussing what are AWS, it is worth noting that the prevalence of misconceptions, often spread by the media, put obstacles in the way of a balanced discussion about AWS. Stigmatized as "killer robots" or referred misleadingly as "lethal autonomous weapons", people may imagine biped androids with guns in their hands. Such terms are misleading since they emphasize the lethal aspect of the system,

which is not a necessary feature of an AWS. For this reason, I prefer to use the term "autonomous weapons systems," as it highlights their autonomous nature, which is more relevant to the topic at hand. In reality, autonomy in weapon systems is a platform-independent functionality: almost every weapon system can be made autonomous, leading to a great variety of devices (Rosert and Sauer 2018). Thus, it is essential to clarify what I mean by "AWS" in this article.

According to a broad definition, an AWS is "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" (U.S. Department of Defense 2012). This definition is a good starting point, but in order to get a more differentiated view – suitable for philosophical discussion – I introduce two distinctions, which hopefully make clear what notion of AWS I am going to discuss herein.

It is necessary here to invoke the distinction from robotics between *general AI* and *modular AI*, as these categories can determine the scope of the relevant issues concerning AWS. Machines equipped with general AI can apply their software to solve, in principle, any problem; essentially, they are general-purpose problem-solvers, like the human brain. Machines like Alphabet's DeepMind are designed to have general AI. By contrast, modular AI is created to excel at a specific task. For example, IBM's Deep Blue is an intelligent chess machine capable of beating human chess-masters, but it cannot be used to accomplish other tasks (Sánchez and Herrero 2018).

Applying this distinction to AWS, a weapon system equipped with general AI could fulfil the public android-like image of a killer-machine, including moving, targeting, and deciding on operations autonomously. For example, the Terminator in the movies has general AI, and as such, these machines would rightfully raise the Hollywoodian fear that they could rebel and turn against humanity (Asaro 2008).

However, I think this worry is ill-founded. Philosophers like Robert Sparrow (2007) take the possibility of these machines seriously, counting on them as intelligent creatures, but the truth is that even if in the distant future AWS with general AI were possible, as far as is known, militaries are not aiming to develop them. Armies anticipating AWS might do so because these machines would lack emotions and other irrelevant considerations during operation, as they would not need skills that are not directly relevant for them to carry out the order given by the commander. Indeed, an AWS with general AI would rather be a drawback than an advantage (Schulzke 2012).

Besides the general categorization of AI, there is a taxonomy of weapons systems concerning their level of autonomy. Marra and McNeil (2013) introduced the distinction between three different kinds of weapons systems by the requirement of the human operator in the decision process. They refer to this process as the "loop". First, those systems in which there is a human "in the loop", i.e. an operator to execute at least one of the processes that is needed; for example, to launch a bomb from an airplane. Second, a system where a human operator is "on the loop", meaning that the system is capable of executing all of the tasks alone but with a human supervisor having veto power; for example, some current on-the-loop systems are the

Phalanx CIWS and the IAI Harop. Third, weapon systems that are fully autonomous, where all human operators are "out-of-the-loop". Throughout this paper, the term "AWS" will refer to out-of-the-loop systems, which are yet to be developed.

2. Jus ad bellum

In JWT, *jus ad bellum* contains a set of principles, prescribing the just use of force and legitimate reasons for going to war. According to Walzer's original account, only self-defensive wars can have a just cause, such as to protect the sovereignty of the state, innocent life and basic human rights (Walzer 1977). Humanitarian intervention – for example, to prevent a genocide – might also be considered as a legitimate initiative. Once a state has entered a war, its ultimate aim has to be the restoration of peace. While the belligerents[1] carry out this aim, the soldiers, as well as their commanders, have to have the right intention. This last rule, even though it is debatable, will be crucial as the source of one central objection to AWS that I discuss at the end of this paper.

2.1. The lower threshold to engage war

The main objection against the development of AWS ad bellum is that their deployment will lead to radical asymmetry in warfare: a military with AWS gains a significant advantage over potential enemies. It is then more likely that such advantaged armies will engage in wars, and hence there will be more – probably unjust – wars in the future. This willingness to go to war can have many reasons, the lower costs of warfare, the reduced risk of losing human lives, and, of course, the higher chance to overcome an enemy (Johansson 2011).

Earlier, Bradley Jay Strawser defended uninhabited aerial vehicle (UAV) technologies against this objection, and I think his considerations can also be used to defend AWS. According to Strawser (2010, 359), "the scope of this issue far exceeds UAVs [...], of course, but strikes at any asymmetry in military-technological development whatsoever." Recall history's great military-technological revolutions: the discovery of gunpowder and the cannon in the 15th century, the steel and steam revolution in the 19th century, and the appearance of nuclear weapons in the mid-20th century (Sánchez and Herrero 2018; Krause 1992). All of these changes resulted in temporary radical asymmetry between belligerents, which could easily cause a series of unnecessary and unjust wars.

If we think that this objection stands, and asymmetric warfare almost always caused unjust wars in the past, which they will also cause in the future, then it is reasonable to extend the objection. Those who reject AWS on this ground should argue that states have to stop developing *any* kinds of new weapons and military technologies – including defensive ones – except for those explicitly developed for

[1] A belligerent is a nation or person engaged in war or conflict, as recognized by international law.

medical purposes. So, UAVs or AWS are not unique in this respect, and one should either argue against all military technologies or anticipate with a neutral or positive outlook all new developments, including AWS.

3. Jus in bello

JWT is more than an ethical theory, as its jus in bello principles are the foundations of International Humanitarian Law. It has two main principles that are essential for the present discussion. The first one is the principle of discrimination: civilians and persons who are *hors de combat* should always be protected by making a clear distinction between combatants and non-combatants. The former group can be legitimately targeted and engaged by enemies using lethal force, while the latter have immunity until they engage in hostile activities. The second principle is the principle of proportionality: while a military accomplishes an objective, it has to resort only to the force necessary in order to limit unnecessary suffering in war.

3.1. Malfunctioning, hacking, stealing

Every machine can potentially be exposed to malfunctioning, stealing, hacking or other security issues during operation. I will treat these different threats as one because their consequences are similar. While developers are usually willing to admit that nothing is flawless, when it comes to AWS, the potential damage can be extensive. Also, even carefully written software cannot prevent the stealing of the physical device (Lucas 2013).

I want to address these issues by pointing out that these risks are present quite generally. Such accidental harm could also affect automated medical diagnosis systems, smart cities, Internet of Things (IoT) devices[2] (…), or autonomous vehicles (AV), but few would argue that by virtue of these dangers the development of these useful devices should be suspended.

First, note that AWS would operate in militarized zones only; hence, it is highly probable that they would not endanger civilians. Second, despite the theoretical possibility, AWS would not be equipped with weapons of mass destruction. These considerations point towards the conclusion that if a problem were to occur, the AWS would cause damage only to the combatants in a battlefield.

Without clear benefits[3], this damage would also be unacceptable, but compared with the other technologies mentioned above, one might say that AWS is in a better position, as a failure in an AV or in a highly complex smart city might also threat-

[2] This includes a wide variety of things, including "a person with a heart monitor implant, a farm animal with a biochip transponder, an automobile that has built-in sensors to alert the driver when the tyre pressure is low, or any other natural or man-made object that can be assigned an IP address and is able to transfer data over a network" (Rouse 2020).

[3] As Strawser (2010) points out, AWS have several benefits for the militaries and can actually prevent the death of human combatants.

en the lives of civilians. A malfunctioning AV can kill innocent people, but this fact rarely leads to arguments against such vehicles. Philosophers instead try to settle the issue of responsibility after the accident has happened.

One could argue that the comparison between AWS and AVs is misleading because the former can cause more significant harm, while an AV would kill fewer people. Note that this utilitarian consideration misses the possible scenario where the car hits a crowd, injuring possibly dozens of people. However, more importantly, it ignores the AV's planned integration with smart cities, which guarantees vulnerabilities, and new, unforeseeable threats.

Every new technology brings up new security issues. However, the trade-off between the anticipated risks and benefits is a more important question. AVs promise to make everyday life so much easier that most of us are willing to view the risks as part of an acceptable trade-off. What has not been acknowledged yet are the similarly significant benefits of AWS for militaries during armed conflicts.

3.2. Discrimination and proportionality

Besides accidents and contingent issues, like malfunctioning or hacking, Human Rights Watch have argued that AWS could not observe JWT's two central *in bello* principles (Human Rights Watch 2012), and they proposed a pre-emptive ban on such machines.

According to their report, AWS will lack the sensory, computational and interpretative abilities to distinguish between combatants and civilians, which is sometimes also a difficult task for a human being. This is because even if a system can distinguish armed soldiers in uniforms from regular citizens, there are certain circumstances where this is not enough. AWS have to recognize illegitimate targets; for example, persons who are *hors de combat*, surrenderers, guerrilla soldiers – who do not wear uniforms and attack in unusual fashion and places – and unusual targets, like civilians, who directly participate in hostilities, and therefore have lost their immunity. Applying the principle of discrimination can be extremely difficult in these unusual situations.

The principle of proportionality has raised similar objections. Human Rights Watch has expressed doubts that AWS will be able to evaluate the exact measures of necessary force in such complex environments as a modern battlefield. They emphasize the necessarily subjective and context-dependent nature of the skills that are needed to observe the principle of proportionality. So, according to Human Rights Watch, AWS theoretically could not have the capabilities to measure what is the proportionate use of force, and therefore, what is legally allowed under the Law of Armed Conflicts.

Even though AVs and other AI-machines have similar capabilities as AWS, a different approach towards them can be spotted in the literature. It seems that scholars have very different expectations regarding AVs: they assume that these machines will reduce reaction time and will also overcome the weaknesses in human judgment. Since human error causes over 90% of traffic accidents, a significant increase in road safety could be expected (Friedrich 2016).

This optimism is strange as I think it is not easier to write an algorithm that can navigate in an urban environment, in particular, considering that human drivers will still be on roads, than to develop an AI system that is able to appropriately distinguish between combatants and non-combatants. Traffic in a metropolis is highly dynamic, with many unwritten rules and unexpected situations, where human judgment seems indispensable. For example, if there is a police traffic control in place instead of traffic lights, AVs have to be able to recognize the signs given by the officer. Also, human drivers often do not drive precisely by the rules, so AVs have to take into consideration the rule-breakings of the drivers too. It is also to be noted that strictly observing the rules in traffic sometimes can be inefficient and annoying. Can AVs decide that breaking the speed limit by 10% is appropriate or not? Besides, pedestrians might also appear at the most unexpected places where they should not, and the software has to handle these situations. Road constructions, small roads without signs, and the signs of other agents in traffic also have to be considered.

These challenges do not seem more straightforward than to decide who is a civilian and who is a uniformed, armed combatant.[4] Various strategies already exist to enhance the recognition of legitimate targets, e.g. transponders, behaviour analysis, location- and time-limited actions, or multi-sensor analysis. Combining these already existing solutions can offer considerable protection to civilians (Hughes 2014). With respect to the proportionality issue, the chances are bigger for AI to be better than humans. Also, evaluating a proportionate attack is a computational task, where machines abilities are better than humans.

Nevertheless, I admit that these are just speculations. It would be an empirical question whether the introduction of AVs to the roads will lead to increased safety or not. The very same is true though for AWS: their use could reduce unnecessary harm and death in just wars, or could lead to the opposite. What is important to see is the unjustified double standard in favour of AVs and against AWS.

3.3. An allegedly AWS-specific issue presented by Purves, Jenkins and Strawser

All the previous objections have the problem that they are too broad, and accepting them leads to the unacceptability of developing automated technologies across a broad spectrum. Purves, Jenkins and Strawser (2015) recognized this issue, and they tried to construct an argument targeting AWS specifically. Their point is that applying AWS would be similar to deploying psychopaths in the battlefield. They based this claim on the assumption that the decisions of the AWS

[4] I do not deny that there is an epistemological difference between friendly and hostile environments (Sterelny 2003). Participants of a friendly situation are aimed at unambiguity and directness, while the uncertainty of signals, unpredictability, mimicry, and camouflage are common elements of strategy in a battlefield. However, this circumstance does not affect the relevant distinction of combatants and non-combatants as everyone is interested in the clear-cut differentiation in this particular case.

cannot be made for the right reasons.[5] Hereafter, I examine this argument in more detail.

Purves, Jenkins and Strawser's strategy relies on a principle in JWT: "there is a positive requirement to act for the right reasons in deciding matters of life and death" (2015). *Ad bellum*, it is not enough to have a just cause to go to war, but the belligerents have to act correctly for this very reason (see Section 2). AWS do not fall under this rule because they do not have the right to enter into wars: politicians and military advisors play the decisive role in these matters.

Purves, Jenkins and Strawser (2015) were not satisfied with this narrow requirement. The authors claim that it can be extended to the whole time of the war, creating an *in bello* obligation for the soldiers on the battlefield to act for the right reasons in their every act. Even though this is not part of the traditional JWT, they refer to authorities who support this idea, like Thomas Nagel (1972) and Peter Asaro (2012).

AWS are, by their very design, not able to act for the right reasons, so they cannot be applied in a just war. To illustrate their point, the authors draw a scenario with a sociopathic soldier, taken to be similar to the application of an AWS:

> Imagine a sociopath who is completely unmoved by the harm he causes to other people. He is not a sadist; he does not derive pleasure from harming others. He simply does not take the fact that an act would harm someone as a reason against performing the act. In other words, he is incapable of acting for moral reasons. It then comes about that the nation-state of which this man is a citizen has a just cause for war: they are defending themselves from invasion by an aggressive, neighbouring state. It so happens that the man joins the army (perhaps due to a love of following orders) and eagerly goes to war, where he proceeds to kill scores of enemy soldiers without any recognition that their suffering is morally bad. He is effective precisely because he is unmoved by the harm that he causes and because he is good at following direct orders. Assume that he abides by the classic jus in bello rules of combatant distinction and proportionality, yet not for moral reasons. No, the sociopathic soldier is able to operate effectively in combat precisely because of his inability to act for moral reasons.

Purves, Jenkins and Strawser (2015) conclude that anyone who thinks that it is problematic that a sociopath soldier would be involved in waging war, has to accept that applying AWS would be just as problematic.

Even though the authors tried to present an argument that is specific to AWS, one can easily make an analogy with AVs in this case: these machines sometimes also have to make decisions about matters of life and death. Consider the widely discussed trolley-type dilemmas arising in the literature (Lin 2016). Purves, Jenkins and Strawser (2015) are aware of this and present a twofold answer against these doubts.

[5] Moreover, the supposedly unfulfilled requirement of acting for the right reasons – or the lack of moral reasoning in general – can be the basis of other deontic objections against AWS; for example, the problem of human dignity (Sharkey 2019) or respect (Skerker et al. 2020). In this paper, I restrict myself only to analyzing the arguments presented in Purves, Jenkins and Strawser's (2015) work.

On the one hand, their response takes into consideration the *raison d'être* of these machines. AVs are created for peaceful purposes in order to make traffic safer, and despite the fact that they sometimes have to make decisions with potentially lethal outcomes, these are not part of their everyday operation. On the other hand, they retort that there is a distinction between the frequencies of the decision-making: while AWS will continuously make lethal decisions as they are supposed to do, AVs will only do so on rare occasions. They consider these two arguments together satisfying enough to differentiate between weaponized and non-weaponized autonomous systems.

It is not easy to argue against Purves and his colleagues' position because they keep it smooth and sophisticated without being too ambitious. At the end of their paper, they position their stance with many qualifications:

> Even if the responses fail to maintain a hard moral distinction between weaponized and non-weaponized AWS, however, we are not ultimately concerned about our argument ruling out driverless cars and other autonomous systems. We ought to meet a high bar before deploying artificial intelligences of any kind that could make morally serious decisions—especially those concerning life and death. It is plausible that no autonomous system could meet this bar (Purves, Jenkins and Strawser 2015).

I agree that we have to be extremely careful before we start to use AVs or wage wars with AWS. However, I maintain that those authors failed to provide a persuasive moral distinction between the two technologies. In the following, I raise three points that do not necessarily falsify their arguments but weaken them considerably.

First, I would like to point out that the conjunction they used to underpin the difference between AVs and AWS contains two different types of conjuncts. The second one – that AVs will make fewer lethal decisions than AWS – is a quantitative argument; however, I think it misses the point. Note that, according to JWT, combatants are legitimate targets, so only the accidental non-combatant killings should count in this comparison. This fact considerably changes the intuitive appeal of the argument, because the difference in this respect is slight. I acknowledge that AWS are still in a worse position, but the boundaries are vague. For this reason, it is difficult to build solid grounds for this objection without telling how many lethal decisions are acceptable in an autonomous machine's life. This task is still ahead of the authors.

The first conjunct – which points out the reason why people make a machine – is a qualitative difference between AVs and AWS, and it is more interesting as I think this embodies the real reason why people are so hostile towards AWS technologies. We do not like things that are designed to take someone's life and, in addition, war and weapons have extremely negative connotations in western culture. These are sometimes legitimate concerns, but given the correct understanding of JWT, most of them are questionable. In a just war, the defensive state has the right to use weapons

no matter what technologies are involved.[6] However, in an unjust war, it does not matter if we use guns, drones or AWS, our actions will not be legitimate. Moreover, in JWT, as one can argue, the purpose of deploying AWS in the battlefield is to protect innocent civilians and to achieve peace faster and with reduced loss, which seem to be desirable goals.

Furthermore, I would like to add that the military chain of command creates special circumstances in moral responsibility, and endows agents with various kinds of ethical status. Due to this fact, we can expect very different things from commanders and soldiers. To highlight the importance of right intention, the authors use Jaworska and Tannenbaum's (2014, 245) example from everyday life:

Consider, first, giving flowers to Mary only in order to cheer her up, as opposed to doing so merely to make Mary's boyfriend jealous. Although the two actions are alike in one respect—both involve giving Mary a gift— the different ends make for a difference in the actions' nature and value. Only the former is acting generously, while the latter is acting spitefully. In one sense, the intended end is extrinsic to the action: one can have and intend an end independently of, and prior to, performing the action, and the action can be described without any reference to the intended end. And yet something extrinsic to an act can nevertheless transform the act from merely giving flowers into the realization of acting generously (or spitefully), which has a distinctive value (or disvalue).

It would be unnecessary to deny that the intentions have a crucial role in our *everyday* moral – and also legal – judgment, but in the military, commanders have the responsibility – and the burden of punishment – instead of their soldiers due to the chain of command (Schulzke 2012). For this reason, it is questionable whether a soldier on the battlefield has to have the right intention in order to morally justify his or actions. This only applies in a special version of JWT, and it is plausible to suppose that it is enough if the commanders give their orders with the right intention.

Finally, I would like to point out that the objection of Purves, Jenkins and Strawser (2015) seems to disregard the intentions of the developers of these machines. We should not be surprised by their analogy between a sociopathic soldier and AWS, because this similarity is intended. As I mentioned in Section 1, military robots will only have modular AI, while lacking moral sense, in order to follow directions precisely. This is how militaries want them to be, so the problematized similarity with a sociopath soldier is not a bug, but a feature.

4. Jus post bellum

Jus post bellum principles aim for a trouble-free transition from war to peace. A significant part of this progress is the accountability of potential war criminals in

[6] Weapons that are below the legally required line of distinction and proportionality, or considered unethical – e.g. weapons of mass destruction, landmines, blinding lasers, expanding bullets – are exceptions. Keep in mind that certain armaments can be made autonomous, but AWS are not weapons themselves (see Section 1).

international courts. Therefore, it is an essential requirement in JWT that every belligerent has to be a liable moral agent. AWS seem to challenge this principle.

4.1. Responsibility gap

According to Robert Sparrow (2007), the main challenge in JWT posed by the deployment of AWS is the so-called responsibility gap issue (cf. Matthias 2004). Suppose that – for some reason – an AWS made a mistake and destroyed a village with civilians only. Note that AWS – unlike a remotely controlled uninhabited aerial vehicle – is targeted and engaged automatically without a human in the loop. Who is responsible for this war crime?[7]

Several possible candidates can bear the responsibility: the most obvious is the AWS itself. Nevertheless, it seems that AWS is just not the right type of entity to bear moral responsibility. It lacks the general AI that would be needed to comprehend the consequences of its actions. Moreover, an AWS cannot be punished in a meaningful way (Sparrow 2007).

Apart from the machine, the programmer might also be liable (Kuflik 1999). This suggestion seems more plausible than the previous one, but still faces serious difficulties. I will mention three of them. First, the codes used in AWS and other highly independent automata are learning algorithms, so the output of a given input is unknown, even to the programmer. Second, programmers usually work in teams, so it would be difficult to identify the one person who wrote the code that led to the tragedy. Additionally, the individual members of the team rarely understand how the full software works in practice. Finally, due to the modular architecture of the system, situations may occur in which developers buy the software from another project to use it in the AWS. For instance, software developed for AVs can be used to automatize infantry fighting vehicles. In this case, the programmers may not know that their codes were used in an AWS; therefore, it would be unjust to blame them.

Finally, one can blame the operation commander who ordered the deployment of the AWS to the battlefield (Lazarski 2002). *Prima facie*, this seems a fair solution, but a similar issue to the previous objection can be raised here. The commander may even have less knowledge than the programmers about how an AWS would respond in different situations. Therefore, it would not be fair to punish him or her for the malfunction of the machine.

Those who object based on the responsibility gap issue argue that – mostly because of the reasons introduced above – nobody can be blamed for the wrongdoings of an AWS. So, according to the *jus post bellum* principles in JWT, their deployment is morally wrong (Sparrow 2007).

Some philosophers think that the existence of the responsibility gap is the ultimate objection against AWS. However, regardless of what improvements can be

[7] Nevertheless, not everyone acknowledges the existence of the so-called responsibility gap issue posed by autonomous technologies (Tigard 2020), but I do not aim to discuss this issue any further in this article. Instead, I focus on the arguments and the debate between those who agree that this is a severe problem.

delivered by AVs and how safe they will be, it is highly probable that they will also make mistakes, even lethal ones – granted though that the number of these events will be insignificant. But, when it happens, the very same responsibility gap is present. According to the literature on the subject, a solution to this issue is not easy.

Parallels with AWS can be made concerning the possible candidates who should bear the responsibility in the case of an accident. These are the vehicle itself, the manufacturer (Gurney 2013) or the owner of the car (Hevelke and Nida-Rümelin 2015). The first option is usually not even considered due to the lack of agency – similarly to the case of AWS. However, the debate between the last two options is lively, because both positions have strong arguments in their favour.

The critical point for the present discussion is the lack of consensus about the responsibility gap issue in the ethics of AVs as well as AWS and the different conclusions that it usually evokes. In the former area of research, scholars appear hopeful about solving the issue, while in the latter case, usually the contrary conclusion is drawn, namely that AWS should be banned as the responsibility gap issue is unresolvable. This double standard, as I have shown, is untenable. We are therefore left without AWS-specific arguments in favour of banning AWS.

5. Summary

Before I summarize this paper's conclusions, I would like to address a general objection against the framework applied through the whole investigation. Some will argue: I should have taken into account that not every war is a just war, or at least it may be difficult to argue beyond any reasonable doubt that it is just. One should not argue for or against AWS on an abstract, idealistic theoretical ground. In light of this, the approach of this article is naive, unrealistic or even unethical. I am aware of this issue, but I would like to add three considerations that may weaken its strength.

First, this kind of objection seems to support my thesis about the unbalanced approach towards the topic. When one starts to read the literature about AWS, one rarely sees these arguments explicitly discussed, neither pro nor contra. This does not mean that concrete, contextual arguments are invalid or unintuitive, just that they are usually not frequently mentioned problems against AWS. In fact, most scholars who oppose AWS construct their arguments in an abstract theoretical space, similar to the one I have used here. This is why I think that it is rational to examine the discussion in an allegedly naive or idealistic approach. I would like to emphasize that, apparently, no one is bothered by this methodology as long as it supports the arguments against AWS, but when the very same approach leads to the contrary, somehow the framework becomes "idealistic".

Second, it is worth reconsidering the goals of developing an ethical theory or policy-making. Generally speaking, in these investigations, one aims to create certain imperatives that people should follow to act virtuously or at least lawfully. Anyone who has the ambition to work out normative theories should suppose that people will follow the rules voluntarily, or the state will have the necessary resource to

force its citizens to follow them. For example, it would be a strange line of thought to legalize thievery because we live in a world where people often break private property laws. In the same way, it is questionable to argue against the JWT framework on the basis that there have been numerous unjust wars in history. In normative projects like considering the ethical status of AWS, we do not aim to describe the facts of the world; instead, we propose certain principles or rules presupposing that these will be observed.

Third, to argue for the ban of AWS on the basis that states are not going to observe the relevant regulations during the deployment of the machines is in a way self-defeating. What guarantees that any law that prohibits AWS will be followed? I think nobody can warrant it, and there are cases indeed when outlawed weapons of mass destruction – like sarin or the VX nerve agent – were used despite their international ban (Murphy 2013; Zurer 1998). Nevertheless, no one would use these unfortunate incidents as evidence that the ban was a wrong decision. Similarly, the potential threat of disobeying the laws and regulations on the use of AWS can hardly be an argument against their deployment. Recalling the previous point, in normative investigations like the discussion of AWS, we have to presuppose that people are generally rule-following; otherwise, all our efforts will be futile.

Despite the above considerations, this paper aimed not to argue in favour of AWS, only to provide a meta-analysis of the debate by pointing out specific biases. By being aware of these partialities, scholars can develop better arguments for or against AWS in the future. Throughout this article, I examined five objections against the deployment of AWS and tried to show that most objections are general to AI-technologies altogether, and Purves, Jenkins and Strawser's deliberately specific argument faces problems. Therefore, without any morally relevant distinction, we should either anticipate AWS with other potentially lethal AI-technologies or rule out all of them.

The first objection was the worry that a military that owns AWS will likely go to war more frequently, because of the reduced costs, and its guaranteed advantage over other militaries. I argued that this applies to every improvement in military technology. So, we either accept that every new development should be banned or accept AWS as just another step in the evolution of weapons systems.

According to the second, contingent objection, AWS will not be able to discriminate between combatants and non-combatants properly and will lack the capabilities to measure the proportionate means of attack accurately. I argued that parallel doubts against AVs could be raised – but in fact, are rarely addressed in the philosophical literature. Therefore, we either ban the development of AWS along with AVs or accept the fact that these technologies can potentially overcome human weaknesses, and thus we should anticipate their development.

The third objection was the worry that AWS could be stolen, could malfunction or could be hacked and these outcomes could lead to disastrous events. I argued that AVs and their supporting technologies, like smart cities or power plants connected to the network, are similarly vulnerable in this respect; in fact, as civilian systems – where security is not always the priority –, perhaps even more so. A smart city has as much vulnerability as an AWS and the consequences of a possible attack or malfunction is also catastrophic.

Perhaps the most challenging issue concerning AI-technologies is the responsibility gap: from intelligent elevator systems through AVs to AWS, many technologies are affected by this (Matthias 2004). However, depending on the type of technology in question, ethicists assess the problem differently. In the field of war ethics, the responsibility gap is usually an ultimate reason to ban AWS in the future. But when it comes to AV, the attitude of philosophers is markedly different: they try to resolve the problem in order to remove the barrier to AV application. I argued that this double standard is mistaken because such an objection rules out a broad spectrum of AI-technologies, AVs among them.

Finally, Purves, Jenkins and Strawser's objection was created specifically against AWS, but its success is debatable. I called attention to two points regarding their insights. First, the military chain of command creates a special context concerning responsibility attribution, so the major purpose of creating AWS is to eliminate unnecessary human emotions and intentions, but the authors have not taken into consideration this fact.

Those who would like to argue against the deployment of AWS have to emphasize its distinguishing characteristic that other AI-technologies or weapons lack. This characteristic can be the basis of a forthcoming argument against them. However, in most of the objections, this characteristic is omitted, which makes the argument too broad to be effective. Purves, Jenkins and Strawser (2015) point out this mistake and attempt to create a specific objection. They succeeded in outlining the distinguishing characteristic – the lack of right intention – but their argument can be challenged because right intention *in bello* for soldiers is not necessary to wage a just war – this requirement only applies to politicians and military leaders *ad bellum*.

References

Asaro, Peter M. "How Just a Robot War Could Be?" In *Current Issues in Computing and Philosophy*, edited by Adam Briggle, Katinka Waelbers, and Philip Brey, 50–64. Amsterdam: IOS Press, 2008.

Asaro, Peter M. "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making." *International Review* of the Red Cross 94, no. 886 (2012): 687–709. https://doi.org/10.1017/S1816383112000768

Friedrich, Bernhard. "The Effect of Autonomous Vehicles on Traffic." In *Autonomous Driving*, edited by Barbara Lenz, Markus Mauer, and J Christian Gerdes, 317–34. Berlin: Springer, 2016.

Hevelke, Alexander, and Julian Nida-Rümelin. "Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis." *Science and Engineering Ethics* 21, no. 3 (2015): 619–630. https://doi.org/10.1007/s11948-014-9565-5

Hughes, Joshua. "Could Autonomous Weapons Systems Be Used Legally Under the Law of Armed Conflict?" Last modified 2014. https://www.academia.edu/8193381/Could_autonomous_weapons_systems_be_used_legally_under_the_Law_of_Armed_Conflict

Human Rights Watch. *Losing Humanity: The Case against Killer Robots*. New York: *Human Rights Watch and International Human Rights Clinic,* 2012.

Jaworska, Agnieszka, and Julie Tannenbaum. "Person-rearing relationships as a key to higher moral status." *Ethics* 124, no. 2 (2014): 242–271. https://doi.org/10.1086/673431

Johansson, Linda. "Is it morally right to use unmanned aerial vehicles (UAVs) in war?" *Philosophy & Technology* 24, no. 3 (2011): 279–291. https://doi.org/10.1007/s13347-011-0033-8

Kahn, Jeremy. "Air Force A.I. Test Raises Concerns Over Killer Robots." *Fortune.* Last modified: December 2020. https://fortune.com/2020/12/21/killer-robots-ai-us-air-force-experiment-u2-spy-plane-artumu/

Kessel, Jonah M. "Killer Robots Aren't Regulated. Yet." *The New York Times.* Last modified: December 2019. https://www.nytimes.com/2019/12/13/technology/autonomous-weapons-video.html

Kuflik, Arthur. "Computers in control: Rational transfer of authority or irresponsible abdication of autonomy?" *Ethics and Information Technology* 1, no. 3 (1999): 173–184. https://doi.org/10.1023/A:1010087500508

Lazarski, Anthony J. "Legal Implications of the Uninhabited Combat Aerial Vehicle." *Aerospace Power Journal* 16, no. 2 (2002): 74–83.

Lin, Patrick. "Why Ethics Matters for Autonomous Cars." In *Autonomous Driving*, edited by Barbara Lenz, Markus Mauer, and J Christian Gerdes, 69–85. Berlin: Springer, 2016.

Lucas, George Jr. "Engineering, Ethics & Industry: The Moral Challenges of Lethal Autonomy." In *Killing by Remote Control: The Ethics of an Unmanned Military*, edited by Bradley-Jay Strawser, 221–29. New York: Oxford University Press, 2013.

Marra, William, and Sonia McNeil. "Understanding 'The Loop': Regulating the Next Generation of War Machines." *Harvard Journal of Law and Public Policy* 36, no. 3 (2013): 1139–1185. https://dx.doi.org/10.2139/ssrn.2043131

Matthias, Andreas. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology* 6 (2004): 175–183. https://doi.org/10.1007/s10676-004-3422-1

Murphy, Joe. "Cameron: British scientists have proof deadly sarin gas was used in chemical weapons attack." *Independent.* Last modified: September 2013. https://www.independent.co.uk/news/world/middle-east/cameron-british-scientists-have-proof-deadly-sarin-gas-was-used-chemical-weapons-attack-8800528.html

Nagel, Thomas. "War and Massacre." *Philosophy of Public Affairs* 1, no. 2 (1972): 123–144. https://doi.org/10.1177/000276427201500678

Purves, Duncan, Ryan Jenkins, and Bradley J. Strawser. "Autonomous Machines, Moral Judgement, and Acting for the Right Reasons." *Ethical Theory and Moral Practice* 18, no. 4 (2015): 851–872. https://doi.org/10.1007/s10677-015-9563-y

Rosert, Elvira, and Frank Sauer. “Perspectives for Regulating Lethal Autonomous Weapons at the CCW: A Comparative Analysis of Blinding Lasers, Landmines, and LAWS.” Last modified: 2018. https://www.academia.edu/36768452/Perspectives_for_Regulating_Lethal_Autonomous_WeapoWe_at_the_CCW_A_Comparative_Analysis_of_Blinding_Lasers_Landmines_and_LAWS

Rouse, Margaret. “Internet of Things.” *IoT Agenda.* Last modified: February 2020. https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT

Sánchez-Herrero, Virginia Romero. “The Ethics of Strategic Artificial Intelligence: An Assessment of Autonomous Weapons Systems through the Just War Tradition.” Last modified 2018. https://www.academia.edu/36382896/The_Ethics_of_Strategic_Artificial_Intelligence_An_Assessment_of_Autonomous_Weapons_Systems_through_the_Just_War_Tradition

Scharre, Paul. “Are AI-Powered Killer Robots Inevitable?” *Wired.* Last modified: May 2020. https://www.wired.com/story/artificial-intelligence-military-robots/

Schulzke, Marcus. “Autonomous Weapons and Distributed Responsibility.” *Philosophy & Technology* 26, no. 2 (2012): 203–219. https://doi.org/10.1007/s13347-012-0089-0

Sharkey, Amanda. “Autonomous Weapons Systems, Killer Robots and Human Dignity.” *Ethics and Information Technology* 21 (2019): 75–87. https://doi.org/10.1007/s10676-018-9494-0

Skerker, Michael, Duncan Purves, and Ryan Jenkins. “Autonomous Weapons Systems and the Moral Equality of Combatants.” Ethics and Information Technology 22 (2020): 197–209. https://doi.org/10.1007/s10676-020-09528-0

Sparrow, Robert. “Killer Robots.” *Journal of Applied Philosophy* 24, no. 1 (2007): 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Sterelny, Kim. *Thought in a Hostile World: The Evolution of Human Cognition.* Oxford: Blackwell, 2003.

Strawser, Bradley Jay. “Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles.” *Journal of Military Ethics* 9, no. 4 (2010): 342–368. https://doi.org/10.1080/15027570.2010.536403

Tigard, Daniel W. “There Is No Techno-Responsibility Gap.” *Philosophy & Technology* (2020). https://doi.org/10.1007/s13347-020-00414-7

U.S. Department of Defense. DIRECTIVE NUMBER 3000.09, November 21, 2012.

Walzer, Michael. *Just and Unjust Wars.* New York: Basic Books, 1977.

Zurer, Pamela. “Japanese cult used VX to slay member.” *Chemical & Engineering News* 76, no. 35 (1998): 7. https://doi.org/10.1021/cen-v076n035.p007