Journal of Military Ethics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/smil20

The Problem with Killer Robots

Nathan Gabriel Wood

To cite this article: Nathan Gabriel Wood (2020) The Problem with Killer Robots, Journal of Military Ethics, 19:3, 220-240, DOI: 10.1080/15027570.2020.1849966

To link to this article: https://doi.org/10.1080/15027570.2020.1849966

Published online: 14 Dec 2020.

Submit your article to this journal

Article views: 340

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=smil20

JOURNAL OF MILITARY ETHICS
2020, VOL. 19, NO. 3, 220–240
https://doi.org/10.1080/15027570.2020.1849966

Routledge
Taylor & Francis Group

Check for updates

The Problem with Killer Robots

Nathan Gabriel Wood

Research Foundation – Flanders, Ghent University, Gent, Belgium

ABSTRACT
Warfare is becoming increasingly automated, from automatic missile defense systems to micro-UAVs (WASPs) that can maneuver through urban environments with ease, and each advance brings with it ethical questions in need of resolving. Proponents of lethal autonomous weapons systems (LAWS) provide varied arguments in their favor, ranging from claims that LAWS will be more effective to arguments that they will be more moral warfighters than flesh-and-blood soldiers. However, the arguments only point in favor of autonomous weapons systems, failing to demonstrate why such systems should be *lethal*. In this paper I argue that if one grants the proponents' points in favor of LAWS, then, contrary to what might be expected, this leads to the conclusion that it would be both immoral and illegal to deploy *lethal* autonomous weapons, because the features that speak in favor of LAWS also undermine the need for them to be programmed to take lives. In particular, I argue that such systems, if lethal, would violate the moral and legal principle of necessity, which forbids the use of weapons that impose superfluous injury or unnecessary harm. I conclude by highlighting that the argument is not against autonomous weapons *per se*, but only against *lethal* autonomous weapons.

KEYWORDS
Autonomous weapons; necessity; non-lethal weapons; just war theory

1. Introduction

Whether we like it or not, the development and deployment of increasingly autonomous weapons systems is the order of the day for advanced military powers that find themselves in conflict. From so-called "smart" munitions, which can identify and track targets, to drones capable of operating with little to no human control, unmanned weaponry is becoming a commonplace on the battlefield. There are many reasons for this growing trend, but perhaps the most compelling (and certainly the most morally salient) reason given is that autonomous weapons, with their ability to better discern legitimate military targets from non-combatants and limit collateral damage, present a possibility to wage wars that involve far fewer (perhaps even no) civilian casualties, thus eliminating one of the most troubling aspects of warfare. In addition, the increased targeting abilities of autonomous weapons allow them to better engage legitimate targets, increasing the likelihood of mission success while simultaneously removing the risk to soldiers who would

CONTACT Nathan Gabriel Wood Nathan.Wood@UGent.be Ghent University, Room 120.065, Blandijnberg 2, Gent 9000, Belgium
© 2020 Informa UK Limited, trading as Taylor & Francis Group

otherwise be carrying out said missions. Put simply, autonomous weapons present a means to potentially wage war in a fashion that is far more discriminate and limited than ever before.

However, despite this advantage, there is serious opposition to the use of what are known as "lethal autonomous weapons systems" (hereafter LAWS), or more colloquially, "killer robots". The arguments and objections take many forms, ranging from fears that faulty or inadequate algorithms will lead to widespread mis-targeting of innocents (Guarini and Bello 2012; Sparrow 2016), to more principled concerns that the use of LAWS will violate the dignity of those humans targeted by such systems (Johnson and Axinn 2013; Purves, Jenkins, and Strawser 2015; Sparrow 2016), or will make it difficult or impossible to rightly determine who should be held accountable when mistakes are made (Sparrow 2007; Roff 2013).[1] These objections, however, may remain unconvincing to proponents of LAWS, as the first points to a mere technical issue, (and one which can plausibly be dealt with or at least seriously mitigated given the rate advancements are being made in facial recognition software and biometric analysis), the second relies on particular intuitions about morality and justice which may not be universally shared, and the last has been shown to be subject to compelling counterarguments (Lokhorst and van den Hoven 2012; Robillard 2017). Yet there exists another argument against the use of LAWS, namely that the best reasons *for* using them simultaneously provide compelling reasons *against* using them.

In this article I argue that taken together, the points in favor of LAWS undermine the overall justification for them. In particular, I argue that the merits of autonomous weapons systems (hereafter AWS) argue against such systems being made to be *lethal*. Thus, I do not argue that the use of robotics in war is problematic, (quite the opposite, in fact) but rather that the use of *lethal* robotics is problematic. Moreover, I argue that given the capabilities claimed of current and future robotics, and the autonomous weapons systems we might create, allowing these to take lives would likely constitute a violation of the moral and legal principle of *necessity* in war, which prohibits the use of any weapons or methods of warfare "of a nature to cause superfluous injury or unnecessary suffering" (Additional Protocol I to the Geneva Conventions, Art. 35.2). I further show that even if such weapons would not violate necessity, international law requires that states demonstrate this before deploying such weapons, and that therefore LAWS will be illegal (at least) until adequate research and testing has been conducted to ensure that these systems are fully compliant with the laws of armed conflict.

The article is structured as follows. First, I cover some preliminary points concerning autonomy and the nature of AWS. With these in place, I then present the strongest argument in defense of LAWS. I then argue that this defense fails, and in fact provides a compelling objection to the use of *lethal* autonomous weapons systems, because such systems would violate the moral and legal principle of *necessity*. With this in place, I then consider objections to my argument and I clarify that the argument is not against the use of autonomous weapons systems *per se*, but only to ones which are *lethal*. Finally, I conclude by briefly presenting a defense of non-lethal AWS.

2. Preliminary points

Before moving onto the arguments, let me clarify two important points regarding autonomous weapons systems. Firstly, which concept of "autonomy" is at stake in *lethal autonomous weapons systems*? Autonomy is a rich concept developed in extensive detail in the philosophical literature, and to attempt to explore that fully here is far beyond the scope of the present work. However, a simple notion of autonomy will suffice for our purposes. Following the definition given by the United States Department of Defense, whenever we say a weapons system is autonomous (which may be thought of as shorthand for "fully autonomous"), what is meant is that the weapons system can, without any contemporaneous human input, engage a potential target based on its own sensory data and internal information processing systems. If, however, a weapons system requires human input at any time during engagement, then it is at most semi-autonomous, as decisions cannot be reached based on the system's capabilities alone, instead requiring a human operator to contribute to some portion of the decision process (U.S. Department of Defense Directive 2017, 13). Notable examples of what, according to this definition, would be (fully) autonomous weapons systems include the point-defense CIWS gun turrets used for missile defense by many navies around the world. These systems can automatically detect and engage incoming missiles or high-speed aircraft without any human input. On the other hand, current drones like the MQ-1 Predator or MQ-9 Reaper represent, at most, semi-autonomous systems, as they are incapable of making targeting decisions without a human operator who is "in-the-loop". In what follows, I will only be concerned with fully autonomous weapons systems, which would be able to identify and engage targets without immediate human oversight.[2]

The second point worth clarifying concerns what autonomous weapons systems are. In some of the literature, AWS are discussed using the language of "robot soldiers" (Anderson and Waxman 2012) or "robot warriors" (Sparrow 2007), with some authors going so far as to maintain that AWS are to be seen "as not simply weapons but a class of combatants" (Roff 2013, 352). This, in turn, leads to discussions which conflate the principles governing weapons development and deployment with the principles governing combatant behavior, to the point that scholars may begin to investigate the moral and legal requirements an AWS must satisfy by first asking what we would require of a human soldier performing the same function (e.g. Anderson and Waxman 2012, 10). Although this conflation is somewhat understandable, given the language used, it is harmful to the debate being carried out, as there are important distinctions between the laws governing combatant behavior and those governing weapons development and use.

In the following, I consider AWS as nothing more than what the United States Department of Defense stipulates them to be, namely, weapons systems and not soldiers. It may be the case that at some point in the (arguably far) future, AWS may become more agent-like, and thus be more properly assessed using the laws which govern combatant behavior. However, current and near-future designs are nowhere near this level of sophistication. To be sure, as weapons, they are highly advanced, but they remain weapons nonetheless. To borrow a phrase from a prominent scholar in the field, "[t]he autonomous robots being discussed for military applications are closer in operation to your washing machine than to a science fiction *Terminator*" (Sharkey 2010, 376). Given

this, the arguments to come will focus entirely on the current moral and legal restrictions on weapons. Of course, it is possible that more agent-like machines may be constructed in the future, and the ethical questions surrounding them are many, but such debates go beyond the discussion here.

With these preliminary points dispensed with, let us now move to the arguments in favor of LAWS.

3. In defense of LAWS

Imagine an autonomous weapon which, in a fraction of a second, can scan a battlefield, pick out various potential targets, clearly determine whether or not they are armed, threatening, or dangerous, and then compare the individuals' features and likely identities against a compiled list of known enemy soldiers or insurgents. With this information in hand, the weapon is then able to rapidly engage the individuals it determines are legitimate targets, and using its superior aiming abilities ensure that all munitions expended hit those targets without endangering anyone else. Before the first shell casing hits the pavement, all enemy contacts are neutralized, without a single civilian being harmed and without destruction of any property or risk to other friendly soldiers.

The scenario presented is clearly fanciful, more the stuff of science fiction than real life, but the underlying message – that autonomous weapons might wage a more discriminate and less bloody form of war – is clear. Moreover, that message is precisely what proponents of LAWS would have us envision; a situation where armed conflict can be carried out with the utmost precision and accuracy, thereby ensuring that no innocents are harmed, while simultaneously keeping our own servicemen and -women free from risk of injury or death.[3] In this way, LAWS would make war a rather different enterprise than it is today, without all the bloodshed and loss of innocent life we have become so inured to. When an autonomous weapons system enters a field of battle, it does not need to preserve its own (for lack of a better term) "life", nor will it respond based on fear or anger. Coldly and with utmost care, it will do its job, eliminating enemy combatants while guaranteeing that no one is caught in a crossfire.

In fact, this lack of emotion and sensitivity is precisely what some of LAWS' staunchest advocates maintain is their greatest strength. Ronald Arkin, a prominent roboticist working on military autonomous and unmanned weapons systems, argues that although there would likely still be errors and innocent deaths on a battlefield dominated by robots, on the whole these incidents would occur far less often than is currently the case, and moreover occur only due to genuine mistakes, instead of being intentional acts of revenge or simple disrespect or negligence, as they too regularly seem to be (Arkin 2009; Arkin 2010). For example, a report from the United States Surgeon General (U.S. Surgeon General 2006) assessing the battlefield ethics of soldiers and marines deployed during Operation Iraqi Freedom found that only 47 percent of soldiers and 38 percent of marines thought that non-combatants should be treated with dignity and respect; roughly 10 percent of soldiers and marines admitted to having damaged or destroyed property when it was not necessary; 7 percent of marines and 4 percent of soldiers admitted to having harmed non-combatants when it was not necessary; and only 55 percent of soldiers and 40 percent of marines said they would report a fellow unit member for injuring or killing an innocent non-combatant; even fewer said they

would report theft from non-combatants, mistreatment of non-combatants, or unnecessary destruction of non-combatants' property. Moreover, the report found that "[f]or all the behaviors under study, [s]oldiers and [m]arines who had high levels of anger were twice as likely to engage in unethical behaviors on the battlefield compared to those [s]oldiers and [m]arines who had low levels of anger" (U.S. Surgeon General 2006, 38), and unnecessary physical violence towards non-combatants was 7 times more likely when the soldier was angry. Roughly equivalent trends were found for soldiers and marines who screened positive for mental health problems (anxiety, depression, or acute stress) or for those who had a casualty within their unit, with unethical behavior being between 1.5 and 2 times more likely given these factors.

Given all of these considerations, the proponent of LAWS can rightly indicate that at least a fair amount of the unnecessary (not to mention illegal) harming of non-combatants and destruction of property is due not to battlefield mistakes or faulty intelligence, but rather to very human reactions to stress, fatigue, anger, and sentiments of revenge. A robot, without any of these physical and emotional responses, would be able to act based only on evidence and reason, and would therefore be far less prone to harming non-combatants or causing more destruction than necessary.

The LAWS proponents' argument might be made even stronger than this though, for there are three areas where LAWS could be made to not only avoid the uniquely human mistakes of battle (i.e. mistakes due to emotional, physical, or psychological strain), but also minimize the instances of genuine errors, of missing targets and hitting civilians or intentionally aiming at civilians due to faulty intelligence.

First, this is because machines are (or at least soon will be) better able to aim than humans can (Guetlein 2005); special sensors can with great accuracy determine range, windage, and a target's rate of motion, and with this information then fire a single round to take down the enemy. In this way, we might think of AWS as highly capable snipers, far superior to any human holding a weapon.

Second, computers can process information far faster than a human, allowing a robot to more quickly respond to incoming information that might alter whether or not a potential target is seen as a *legitimate* target or instead a non-combatant. Moreover, a robot will not be as susceptible to confirmation bias, where incoming information is only interpreted in such a way as to substantiate or validate already held beliefs (Nickerson 1998), making it more likely to fairly judge all evidence at its disposal.[4]

Finally, and most importantly, a battlefield robot does not necessarily need to protect itself, which allows it to incur more risks for the sake of being absolutely certain that a particular agent is indeed a legitimate target and not simply a kid playing with a toy gun, or a young man cleaning a firearm but otherwise not engaged in harmful behavior.[5] These types of cases are what many opponents of LAWS point to as problems, scenarios where a computer may falsely identify some person as a combatant when they are in fact holding something that is not at all threatening, or they are indeed armed but are themselves non-threatening. While these cases do indeed point to a potential issue (if LAWS targeted such people, we would have very good grounds to object to their use), the fact is that a LAWS does not need to target anyone unless it is certain that that person is a threat, because there is no moral loss if the LAWS is "killed" due to overly cautious behavior when it comes to targeting decisions.[6] A robot can assess a situation, and if it is unsure, it can quite simply get closer, the better to see exactly what is going on or

better hear what the potential targets are saying. A soldier, on the other hand, cannot act in such a way, or at least cannot be reasonably expected to act in such a way, as every moment of hesitation potentially places him or her at greater risk. Even more clearly, a soldier cannot reasonably be expected to "simply get closer" to potential combatants to be sure whether or not they are indeed combatants, as this makes that soldier an easy target, a moral risk often not outweighed by any significant benefits. And these same points concerning proximity will also hold if an enemy is difficult to target due to some environmental or contextual factors like weather, the presence of civilians nearby, or even the use of human shields; while a flesh-and-blood soldier will have to make a quick decision which takes into account his or her own safety, a robot can act more cautiously with its targeting in order to protect nearby civilians.

To be clear, this is not to say that soldiers do not have to be sure about their targets, or that they may forgo incurring certain risks to increase their confidence in targeting assessments or aim; this is only meant to show that a robot can and should be expected to incur more risks than a soldier *because the preservation of the robot is not morally worth risking harm to innocents*.[7] Given these factors, it is clear that a robot, in addition to not being negatively affected by anger, mental illness, or feelings of revenge, will also be able to, on an everyday basis, make fewer mistakes in combat due to its better aim, better targeting, and ability to take on increased risks for the sake of certainty.[8]

Given all of these factors, the proponent of LAWS would maintain that robots would be not just more effective, but also more ethical than human beings on a battlefield. This is because, according to the LAWS proponent, robots will be capable of being far more discriminate than their human counterparts, and will be able to act with a minimum of force and violence in a given situation, since they would be capable of incurring many more risks for the sake of precision and non-combatant protection. Moreover, a robot will never be subject to the physical, emotional, and psychological strains which contribute to so many breaches of battlefield ethics.

4. Why *lethal* autonomous weapons?

Now, for the sake of argument, let us grant the proponent of LAWS all of the empirical premises up to this point. Let us simply stipulate that robots are (or soon could be) capable of processing information more quickly and accurately than humans, that they can do so without any of the particularly human failings of reasoning (e.g. confirmation bias or letting emotions distort facts and value judgments), that they can aim weapons more accurately than humans, and that they can reasonably be expected to incur more risks than humans in order to be more sure of their targeting decisions and accuracy when engaging enemies. Moreover, let us also grant the LAWS proponent the conclusion reached above, that these factors entail that a war fought by AWS would be more ethical than one waged by flesh-and-blood troops, thereby making it morally preferable to deploy robots instead of humans. The question then is whether this lends any credence to the idea that we should or even may create and deploy autonomous weapons systems which are designed to be *lethal*? I argue that if the LAWS proponent is granted the above points, then the answer is no.

The simple argument for this conclusion is essentially that if autonomous weapons systems possess the capabilities and strengths that their proponents would have us

believe, then there is little reason why they would *need* to take lives, and if there is not a need for them to kill, then it would be morally wrong for them to kill. To see this, let us envision a robotic weapons system of the kind described above and consider it in an actual battlespace.

The robot, we have assumed, can process information rapidly, has aiming abilities better than any human, and can move closer or relocate in order to get a better shot, even if this imposes some added risks to the robot. Now, suppose this robot is deployed to a conventional armed conflict between states, where both parties have uniformed soldiers acting within the legal framework of war. In such a scenario, the aim of both armies is to force the opposing government to capitulate, which is normally achieved by imposing casualties on the opposing army in order to defeat them militarily.[9] Importantly, when we say that an army imposes *casualties*, this does not mean the army necessarily kills (or intends to kill) anyone, as a casualty is any person who is no longer combat-effective, regardless of the reason (U.S. Department of Defense 2018, 32). Moreover, if an army is capable of imposing casualties without killing the enemy, then they are morally and legally required to do this, as it is morally and legally wrong to kill needlessly (Mayer 2013).[10] Even more strongly, there are clear legal prohibitions against the use of weapons that guarantee death when wounding is equally effective for victory, and these prohibitions hold full force with regards to LAWS, which are, at base, nothing more than weapons.[11] So, if the robot can aim better than any soldier and take on extra risks to get cleaner shots, and moreover, if that robot's destruction holds little moral disvalue (at least in comparison to the loss of a human life), then it should impose whatever casualties are necessary for victory (in the strict sense of the term "casualty" defined above) but do so without aiming to kill anyone, as this will contribute most to victory while staying as true as possible to the moral and legal codes of war (more on this latter point below). This means the robot should be shooting soldiers in areas that are less likely to be lethal (e.g. extremities or lower torso), or harming them in ways that will render them combat-ineffective but impose comparatively low risks of death.[12]

Now, if robots are deployed to a non-conventional battlespace such as an asymmetric conflict or a counterinsurgency environment, the situation will be slightly different. First of all, there will be more difficulty in determining who are combatants and who are civilians, due to the lack of clear combat dress for rebel or insurgent fighters. This, however, will make it more important that the robots avoid lethal behavior, because it will be possible (perhaps even likely) that some of the targets engaged are not combatants at all. Given this, there is strong reason to aim by default for non-lethal applications of force.[13] Moreover, rebel and insurgent fighters are usually taken to be illegitimate combatants, meaning that they do not enjoy the same protections that rank and file soldiers in an official military do. In particular, illegitimate combatants are considered to be, in a real sense, criminals, given that they carry out violent, often lethal, attacks without the permission or authorization of a legitimate authority entrusted with coercive powers. Now, given that such fighters are taken to be acting criminally (at least in some respects), one may think that there is less need to minimize the force used against them or reduce the number of rebels/insurgents killed. However, this conclusion is incorrect.

The actions of rebels and insurgents are seen as illegitimate because their fighting is outside the scope of the traditional laws of war, but this means they must be treated in the same way we would treat any other criminals; we are not permitted to simply find

and kill them, but must instead make an effort to apprehend them for the purposes of trying them for their crimes. As an analogy, when a police raid targeting a mafia hideout is planned, the SWAT or armed response units are not told to shoot anyone they see who is armed or threatening, but instead are expected to do their best to arrest all individuals so that they may be given fair trials later.[14] The same is true for rebels and insurgents. If the state chooses to brand their fighting illegitimate and hence illegal, then the state should ideally interact with them as they would with any other criminal element, which is to treat them as innocent until proven guilty. With this comes a requirement that the state put forth a real effort to apprehend the rebels/insurgents and put them on trial, which precludes them being intentionally lethally targeted during military engagements.[15] If the state does not wish to treat them as criminals, perhaps because they would rather lethally target the rebels/insurgents, then they must recognize them as normal combatants (i.e. *not* illegitimate ones), but this simply leads us to the argument above, that the robot should create casualties while aiming to avoid kills.

Thus, when fighting (potentially) illegitimate forces such as rebels or insurgents, AWS should not be lethally targeting individuals, both because there is a higher risk that the targets are actually civilians and thus not liable to any harm, and because any actual combatants are to be treated as criminals and hence arrested and tried for their crimes, which precludes the use of deadly force unless absolutely necessary.[16] And if we assume robots are as good as some proponents of LAWS say they are (aim that is far better than a humans', better than human judgment, ability to take on extra risks, etc.), then there is little reason why they would plausibly need to shoot to kill; as gruesome and painful as it sounds, a bullet through the knee will do perfectly well in keeping an insurgent on the ground until he or she can be arrested by military or police forces, and will not create an unreasonably high risk of death provided the wounded individual receives medical attention in a somewhat timely fashion.[17] And even if the insurgent cannot be apprehended and manages to escape, he or she will almost certainly be, at the very least, combat-ineffective due to his or her wounds.[18] Given these points, it is hard to see what significant military advantage would be gained by designing AWS to make lethal targeting decisions. And without such advantage being present, the added harm due to lethal targeting arguably constitutes a great deal of "unnecessary suffering", in violation of the principle of necessity.[19]

Before moving on, it is worth noting that the arguments of this section are based on general assessments of warfighting scenarios and the usual responses that are justified (both morally and militarily). However, this is not to say that one could not imagine cases where *the only way* to secure some good or avert some evil is through the lethal targeting of enemy combatants (either legitimate or illegitimate).[20] However, even if one imagines a situation where *the only* way to achieve some morally worthy goal is via the killing of enemy combatants, this still does not constitute an argument for LAWS performing the killing, since even in cases where lethality may be necessary, it still must be demonstrated that it is the robot that must make the lethal targeting decisions, and that the robot must be able to do this without any human input. Put differently, there may be times when killing is the only right choice, but for such cases we can have the robot ask a human overseer if lethal targeting is permissible. Then the weapons system would be capable of taking lives, but in cases where it does it would no longer be

fully autonomous, instead becoming momentarily a semi-autonomous system where a human being is "in-the-loop". When lethality is deemed unnecessary, the robot could maintain full autonomy, making targeting decisions and carrying them out without human input, but for cases where it is deemed necessary to kill, the robot could simply defer that decision to a human operator. (This would also go a long way toward alleviating many of the more principled objections to autonomous weapons, such as those presented in Section 1.)

Given this possibility, in order to show that AWS should be made to be *lethal*, one must show both that there are situations where lethality is necessary and that the lethal targeting decisions in those situations cannot or should not be delegated to human controllers.[21] Without such an argument, caution would dictate that we, at most, design robots which are fully autonomous and have the physical ability to kill, but which briefly compromise their autonomy when lethal targeting is considered as an option.

5. LAWS and necessity

Up to this point, the arguments have dealt primarily with moral concerns, but the points developed bring forward a legal problem as well, namely, that deploying lethal AWS would likely violate the legal principle of *necessity*, a core tenet of the international law of armed combat.

Necessity, due to its centrality to the law of armed combat, is given slightly varied formulations in order to capture the various ways in which it is used. For example, there are statements of necessity that govern the weapons and materials that may be used in war (AP (Additional Protocol to the Geneva Conventions) I, Art. 35.2; 1868 Declaration; 1899 Declaration), there are statements that govern the ways soldiers may interact with their enemies (ICRC *Soldier's Rules*; USMC *Law of War* training handout; U.S. *Operational Handbook*), and there are statements which govern how soldiers act with regards to civilians (AP I, Art 57.2.a). However, at the core of all renditions of necessity is the demand that one only impose harms necessary for the achievement of some military end. Conversely, necessity forbids any actions which cause "unnecessary harm, because the same benefits could have been achieved by less harmful means" (Hurka 2008, 128; see also Lazar 2012 and 2014 for similar statements). Originally, the principle was put forward as a purely moral imperative aimed at ensuring that soldiers and states would have a righteous intent when waging wars,[22] and this moral flavor is retained in much of the philosophical literature on ethics and war.[23] However, the underlying intuitions driving the principle have in the last two centuries given rise to a full-fledged legal requirement, codified most prominently in the Additional Protocols to the Geneva Conventions (abbreviated as AP). The most conspicuous statement of necessity can be found in AP I, Art. 35.2,[24] which states that:

> [i]t is prohibited to employ weapons, projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering.[25]

Now, as discussed in Section 2, the first point worth stressing with regards to LAWS and necessity is that autonomous weapons systems, lethal or otherwise, are weapons and *not* soldiers.[26] Thus, the question is not whether the robot makes targeting decisions that

would pass the bar set by necessity for human warfighters in similar situations, but rather with whether the robot, as a weapon, is made such that it is "of a nature to cause superfluous injury or unnecessary suffering". In the same vein, we might alternatively ask whether AWS, if made lethal, would go against the legal principles laid down in treaties such as the 1868 St. Petersburg Declaration banning exploding bullets or the 1899 Hague Declaration 3 banning bullets "which expand or flatten easily in the human body".[27]

In each of these treaties, the aim is to prohibit weapons which cause more suffering than what is necessary for military victory. Moreover, the 1868 Declaration clearly states that it is sufficient for victory to disable men without killing them (in line with the arguments of section 3 above), thus making clear that death itself can be seen as a form of "unnecessary suffering" in cases where victory is possible without killing (Roberts and Guelff 2004, 55). But again, and importantly, the restriction here is on weapons, and not on targeting decisions made by combatants. Thus, nothing about these treaties demands that soldiers only aim for "non-lethal" areas of enemy soldiers, nor that they do their utmost to avoid killing. What the treaties do underpin is a prohibition on weapons which ensure death when a weapon that merely disables is equally effective for securing victory.[28]

With regards to AWS, the same principles would apply, since AWS are, at core, nothing more than weapons. If an autonomous weapons system can secure victory without killing, or if an autonomous weapons system is capable of disabling men without rendering their deaths inevitable, then designing such a system to be lethal would constitute the employment of a weapon "of a nature to cause superfluous injury or unnecessary suffering". And if the arguments of LAWS' proponents presented in section 2 above are granted, there are strong reasons to believe that AWS would be equally effective warfighters without lethally targeting enemy combatants. This is because AWS could be designed to aim only at extremities or the lower half of a target's torso, constraints that are increasingly within the bounds of what autonomous systems can achieve.[29] Such measures would not guarantee that all engaged enemies survived, but it would drastically reduce the number of enemy fatalities without plausibly imposing any significant reduction in combat effectiveness or mission success.

However, one might object that the preceding points, while plausible, are all conditional on non-lethal AWS being as effective as LAWS (i.e. lethal AWS), an assumption which may well not be true. Thus, one might argue that (1) *if* non-lethal AWS could be as effective as lethal AWS *then* this would, on grounds of necessity, constitute an objection to the latter's use, while also maintaining that (2) the antecedent of this conditional simply is not and will not be the case. Thus, the argument would be that there is some military advantage gained by programing lethality into AWS, and that the advantage is significant enough to permit that added harm.

There is no direct response that one could give to this objection. The laws of war (and the generally accepted moral principles governing conduct in war) allow belligerent parties to increase the level of violence, injury, and suffering, so long as that increase is within the bounds of proportionality and serves some military end.[30] Thus, if LAWS were to be more capable fighting machines than non-lethal AWS, then there would not, at least on grounds of necessity, be an objection to their use. However, even given this fact, it still remains the case that nations are not permitted to deploy weapons which impose greater suffering or more serious injury simply because said

nations *believe* that these weapons will contribute more to victory. Rather, nations are required to *ensure* that new weapons will be in accordance with all principles laid down in international treaty, a provision given as a corollary to Art. 35 of AP I and stated explicitly in Art. 36:

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

This means that even supposing there is added value, militarily speaking, for making AWS lethal, it is still the responsibility of states wishing to use these weapons to *first* determine whether or not that is indeed the case, and if so, whether the added value is of a degree significant enough to merit making a non-lethal weapon into a lethal one. This is especially important given that the 1868 Declaration, a treaty which represents the customary international law of necessity, outlaws particular types of weapons specifically because they make death inevitable in cases where disabling is sufficient for victory. Moreover, given that (according to LAWS proponents) AWS have better aiming abilities than human soldiers, will not flinch, hesitate, or withhold fire, and given that they can perform calculations and assess data far faster than a human, it is highly likely that LAWS will in general be *far more lethal* warfighters than flesh-and-blood soldiers (Guetlein 2005, 5), making it all the more evident that states must first demonstrate that such lethality is warranted (or even useful) before deploying these weapons systems. All in all, these factors make clear that even if there is some military advantage to be gained by making AWS lethal, this must be compellingly demonstrated before the use of any such systems is permissible.

To sum up, the points in favor of AWS make it likely that lethality is unnecessary for these systems to be effective. Moreover, as progress is made in robotics and AI, the marginal gains of programing AWS to be lethal will likely decrease, as such systems will be increasingly capable of disabling combatants in ways that are non-lethal. Moreover, given the requirements of AP I, Art. 36, even supposing that there is some military advantage to be gained by making AWS lethal, that advantage must be compellingly shown to be the case before such systems may be employed. In addition, any argument in favor of LAWS must show not just that lethality is (sometimes) militarily advantageous, but also that in such cases the lethal decisions must or should be made by the robot itself and not by some human who is momentarily brought "into-the-loop". Until such arguments and empirical data supporting LAWS are provided, the arguments here presented indicate that these systems would be not just in breach of the moral principle of necessity, but also the legal treaties which codify that principle in international law.

6. Objections

Before concluding, there are three objections that must be dealt with. The first concerns the precise formulation of the conclusion being argued for here, the second deals with a technical point regarding a number of LAWS currently in development, and the third focuses on a legal point one might raise against the arguments so far.[31]

First, one might grant the arguments so far presented, but maintain that they speak in favor of robots being constructed and programed to use a minimum of force, but not necessarily to be entirely non-lethal. Thus, one might grant that robots should perhaps be generally aiming for non-lethal or less-lethal applications of force, while also claiming that there still will exist cases where lethality is strictly speaking necessary, and thus AWS should be able to use lethal force in such cases.

This point is philosophically weaker than the one I am making, and for that reason it is easier to defend theoretically, but this modification of the conclusion makes it far more difficult to employ in practice, since it is often very difficult to determine what indeed counts as the "minimum of force" necessary. In fast-paced combat environments with incomplete and sometimes contradictory information, it is often an astoundingly hard task to determine what the least harmful means is. Moreover, necessity sometimes demands that trade-offs between benefits and harms be considered, an evaluative exercise that is not only factually but also morally complex and nuanced. That is precisely what makes necessity so difficult to rigorously employ in practice. A robot may fare better when assessing information, but judgments of a moral or legal nature will be beyond its abilities, and such judgments are often central in determining what force is permitted, minimal or otherwise. This means that while it may be more theoretically satisfying to argue that LAWS should be built to use a "minimum of force", this will in practice amount to them being made to be non-lethal. Moreover, until AWS are sophisticated enough to be able to competently and reliably determine what force is necessary and minimal, programing them to be able to make lethal decisions will present grave risks that they in fact impose unnecessary or superfluous harms. Thus, until sufficient advances in robotics have been made, a morally and legally cautious outlook would point to making AWS non-lethal.

The second objection one might make is that the arguments so far presented are heavily predicated on the notion that an AWS may be made to be lethal or non-lethal purely in virtue of the targeting decisions it is programed to make. This, in turn, would seem to indicate that the arguments apply only to LAWS that fire bullets or other ordnance that operate purely via the transfer of kinetic energy. However, a large number of LAWS currently in development or under consideration mirror, or are based upon, current unmanned aerial vehicles (UAVs), systems which fire missiles and other explosives. Thus, so the objection goes, lethality will be built in by default, given that one cannot fire a Hellfire missile in a "non-lethal" way.

However, to this objection one may respond using the same line of argumentation presented above in connection with AP I, Art. 36: if we suppose that the firing of explosive-tipped missiles (which we may for the sake of argument assume are lethal)[32] provides a direct military advantage not possible otherwise, it still remains to be shown that an AWS must be able to fire the missiles without any human input. If this is not the case, then there is no added advantage to having lethal fully autonomous UAVs, and so caution (and the preceding arguments) would advise against them.

The objector might then rejoin that part of the justification for making UAVs autonomous is for purposes of air combat defense, as a computer may be better able than a human pilot (remote pilots included) to avoid incoming ordnance and counter enemy maneuvers, adding that part of the computer's strength lies in its speed of assessing information and deciding, a factor which might make the necessity of asking a human

handler's "permission" before firing disadvantageous. But to this, one can easily counter that the autonomous UAV could, on first detecting incoming enemy aircraft or weaponry, simply ask the human's permission to respond with deadly force, thus leaving the human responsible while preserving the advantages of an AWS. And this point may be made more broadly; whenever an AWS encounters a scenario where lethality is expected to be necessary, or even when it is merely seen as a possible necessity, the AWS could be made to request a human's permission to engage lethally. This need not entail that humans make split-second decisions alongside the AWS, as human handlers can give blanket permissions that the AWS go forward employing whatever force it deems appropriate, lethal or otherwise. But by locking lethality to a human's active decision, it becomes possible to eliminate many of the worries surrounding AWS, while still retaining the benefits that such systems might bring. Moreover, in those cases where a human cannot be reached or does not respond quickly enough, the robot will not be incapable of acting. It will merely be incapable of acting in a way that it expects to be lethal. Such caution might indeed lead to deaths of civilians or friendly combatants, but on average a non-lethal AWS will cause far fewer deaths than one which can make normally lethal targeting decisions. The main point remains that there are risks that a lethal autonomous UAV might be in breach of AP I, Art. 35.2, and so by Art. 36, it is the responsibility of states wishing to develop and deploy such weapons to first ensure that they will not in fact be in breach of international law. Until such research is carried out, lethal autonomous UAVs will present a legal problem.

A final response to this objection is that it is not clear that lethal autonomous UAVs similar to current designs would be more militarily effective than competing potentially non-lethal designs. This is because UAVs which fire missiles carry with them a significant operating cost, given the sheer cost of modern missiles. As an example, the AGM-114 Hellfire missile, the standard weapon of the MQ-1 Predator drone, costs roughly $117,000 per unit (U.S. Department of Defense 2016). By comparison, standard ammunition for rifles and machine guns costs, at most, only cents per round. Given this, if micro-UAVs like the WASP[33] were to be outfitted with light arms and developed as AWS, these could operate at significantly reduced cost, allowing for more to be deployed. Moreover, given that such systems can be easily transported and launched by hand, and that they can more easily maneuver through urban environments, these could also provide more effective direct support to ground troops in combat zones. These points indicate that potentially non-lethal UAVs may in fact be militarily superior to lethal UAVs like the Predator or Reaper drones, considerations which call into question whether it would be at all acceptable (morally or legally) to deploy the latter, which inflict greater injury and suffering but which may turn out to be less militarily effective.

The third objection one might raise is that the arguments so far presented fail to appreciate just how permissive the rules of war are. More specifically, one might object that under the laws of war, soldiers are permitted to target enemy combatants simply because of their *status* as combatants, and not because it is *necessary* or *useful* that they be harmed.[34]

To this objection, one must first clarify that while one's status as a combatant does greatly loosen the requirements for being targeted,[35] combatant status is not the only thing necessary to permit the targeting of an individual. Rather, it must be the case that the targeting of an individual actually furthers military endeavors. This is because

the law of war, arguably, is "inspired by the desire to diminish the evils of war by protecting *both combatants and noncombatants* from unnecessary suffering", and to that effect, it "requires that belligerents refrain from employing any kind or degree of violence which is not *actually necessary* for military purposes" (U.S. Army 1956, 3 (emphasis added)).[36] And this restriction does not merely limit violence to that which is *advantageous*, but rather demands that one only impose violence which is *necessary* (Paust 1974; Horton 2006). The reason for this is because a permission to impose any harm which merely provides *some* military advantage would be tantamount to endorsing the principle of *Kriegsraison*, which was expressly repudiated during the Nuremberg Trials.[37]

In addition to these points, the arguments advanced in this article have to do with the principle of necessity as it pertains to weapons development and deployment, and so an objection based on the rules governing combatant behavior are beside the point. More particularly, AWS, *which are by definition weapons systems and not combatants*, are governed by the law of necessity codified in AP I, Art. 35.2, and this principle allows for increased suffering and injury only to the extent that that suffering and injury serves some military purpose. If that is not the case, then the suffering and injury is by definition unnecessary, and hence illegal. This point is further strengthened by the 1868 and 1899 Declarations, documents which are representative of customary international law, and which explicitly prohibit weapons that impose unnecessary suffering. Thus, in order for lethal AWS to be permissibly deployed, it must first be shown to be the case that lethality affords some military advantage not present if the weapons are made to be non-lethal. This is especially the case if the weapons are of such a nature that it is plausibly the case that they could be effective without being lethal.[38]

7. An ethical future for autonomous weapons

So far, it has been argued that the capabilities of AWS undermine the arguments for making such systems lethal, but this should not be taken as an indictment of autonomous weapons in general. Researchers like Ronald Arkin present a compelling case for using robotics in warfare, not just for reasons of effectiveness or protection of one's own troops, but because there are strong reasons to believe that robots will be capable of waging a more moral form of war than regular soldiers ever could, due to the many factors discussed above. To be sure, this is not to say that current technologies place robots at an advantage with respect to moral decision-making, but to indicate that the outcomes they bring about will be more in keeping with the moral and legal principles governing conduct in war.

Given this, though we have strong reasons to object to *lethal* autonomous weapons systems, we ought not impose a blanket ban on research or deployment of autonomous weaponry more broadly. By building better machines and exploring the various ways that moral principles might be codified or reasoned with by a machine, it may become possible to create robotic soldiers whose actions are far more in accordance with morality than those of their human counterparts. Such developments would go a long way to reducing the evils of war, and may even be capable of making warfare an enterprise with little to no deaths on all sides (even when one side's combatants are all human), ideally culminating in a situation where robots disarm combatants and then simply let them go about their business. To use the words of John Canning,

> [o]ur 'dream machine' is one that would confront an enemy combatant on the battlefield, physically remove the rifle from his hands, saw the rifle in half with a diamond-tipped saw, hand the two halves back to him, and then tell him to 'Have a nice day!'. (Canning 2009, 14)

And this brings us to the final point, which concerns a misconception about autonomous weapons. For many, there is an intuitive dread attending any mention of killer robots, with our minds conjuring images of *Terminators*, robots without feeling who hunt you down and kill you for no other reason than that their programing says to do so. However, if robots are not lethal, but instead are designed and deployed to act as Canning's "dream machines", fully adhering to the laws of war and being less violent than human soldiers besides, then it becomes clear just how good autonomous weapons could ultimately become. In regular conflicts, they could quite simply break the enemies' weapons and then leave them be. This would render each soldier combat-ineffective (which is the goal of combat in conventional conflicts) without imposing a single harm on those soldiers. In irregular conflicts, where enemy combatants are deemed illegitimate, the robots could walk into insurgent camps or terrorist hideouts, grab the combatants, and drag them back to the nearest police station kicking and screaming. They would be bruised, but otherwise no worse for wear, and then a proper trial could take place to determine whether their fighting was criminal or an act of legal opposition.[39] Clearly, this picture is likely decades out still (if not further), but the prospect of having war look like this is a chance worth exploring fully. For these reasons, we should welcome the development of autonomous weapons while doing our utmost to ensure that they are programed in such a way as to adhere to *all* laws of war, bearing in mind how their own capabilities will affect the moral and legal prescriptions in a given scenario.

Notes

1. Additional objections to LAWS are developed in Killmister 2008 and Roff 2014, but their critiques are focused on more nuanced aspects of command and control or asymmetric warfare, and not particularly with principled objections to LAWS.
2. For further discussion of the concept of autonomy within the context of military robotics, see Lin, Bekey, and Abney 2008, 103–105; Roff and Danks 2018, 4–5. See, also, Roff forthcoming and Roff 2013, 353–354 for a deeper treatment of the differences between "autonomy" and "autonomous" as these terms pertain to debates on robotics.
3. See, e.g. Arkin 2009; Arkin 2010; Guetlein 2005.
4. It is worth mentioning that this point is subject to exceptions, as robots can still fall prey to human biases as a result of training or programing which is conducted by humans. This, however, should not be taken to indicate that robots necessarily do or must have biases. Instead, it cautions us to be aware of what biases a robot may learn, and, should they in fact be learned, search for ways to eliminate or at least mitigate the negative effects of those biases.
5. Cf. Lin, Bekey, and Abney 2008, 80.
6. This may not be strictly true, as the loss of very expensive machinery does plausibly hold moral disvalue, given that money which could have been used to help people elsewhere will now be used to replace that machine. This, however, is a subtle issue and adequate treatment is far beyond the scope of this article. See Broome 1991 and 2004 for further reading on the moral weighing of goods and lives.

7. Again, there may be moral value to preventing the robot from being captured or destroyed, but that value will (arguably) not exceed the moral value of a human life. For further discussion of this point and others, see Lin, Bekey, and Abney 2008, esp. 79–81.
8. All of these points and more are concisely put forward in Arkin 2010. See esp. 333–334 for an enumeration of the merits of LAWS.
9. It is worth noting that it may be possible to achieve victory in some instances without directly imposing physical harm by the use of weapons on an enemy, for example, if one has diplomatic or economic options which would force an enemy government to accede to reasonable demands. In such cases, these other options are preferable on grounds of necessity, and so violent options should not even be considered.
10. This is the standard approach to warfare codified in international law and encoded into basic codes of conduct for soldiers. See, for example, Additional Protocol I to the Geneva Conventions (hereafter AP I), Art. 35.2 (Roberts and Guelff 2004, 442), the so-called *Soldier's Rules* drawn up by the International Committee of the Red Cross (hereafter, ICRC) and present in the U.S. Operational Handbook (Green 2017, 395–398), or the U.S. Marine Corps' training manual for rules of engagement, which states that "[w]hat we must ensure is that our decisions and actions minimize unnecessary suffering to the enemy as well as to any civilians and noncombatants involved" (U.S. Marine Corps Training Command 2015, 5).
11. The legal prescriptions referred to here are the 1868 St. Petersburg Declaration, and the 1899 Hague Declaration 3, both of which are discussed at greater length in Section 5 below.
12. Importantly, an AWS designed to make "non-lethal" targeting decisions might still cause death in those targeted, as there is always a possibility that a wound becomes fatal for any number of reasons, even if the wound imposed is highly unlikely to be fatal. The main point is not about whether or not the robots actually impose death, but whether they are designed to aim to impose death. For more general concerns about labeling weapons "non-lethal", see Kobayashi and Mellen 2009.
13. The arguments presented in Guarini and Bello 2012 further strengthen this point, as they stress that "in theaters of activity involving mostly noncombatants, differentiating between combatants and noncombatants will often require the appropriate attribution of mental states" (129), something which robots are (at least now) incapable of doing with sufficient reliability. Given this, allowing AWS to make lethal targeting decisions in such environments presents a problem not just for necessity, as I am arguing for here, but also for the principle of discrimination, as an AWS may very well make targeting mistakes in such contexts. This argues forcefully against AWS being made to be lethal, and may even indicate that AWS should use minimal levels of force in these types of environment.
14. Obviously, if a mafia thug engages police units, they are permitted to use force to defend themselves. But even in such a self-defensive situation, they are still morally required to use only as much force as is necessary for self-defense and to act so as to maximize chances of apprehending the thug without inflicting mortal wounds in the process. The same is true of situations where the criminal elements are threatening innocent bystanders; the police may use lethal force to the extent that such force is *necessary* for averting the threat posed by the criminals, but if an equally effective non-lethal option is present, they are required to defer to that option.
15. Coincidentally, FM 3–24, the U.S. Army/Marine Corps manual on counterinsurgency, explicitly and regularly states the importance of reducing violence in general when conducting counterinsurgency, and in particular arresting insurgents and fairly trying them in the existing legal structure(s) in order to establish and/or reinforce rule of law, as this has a profound impact on the success or failure of a counterinsurgency effort. See U.S. Army/U.S. Marine Corps 2014, esp. sections 13–61 to 13–70. See also Wood *forthcoming* for a concise treatment of FM 3–24, which highlights the importance of reducing force and killing for the purposes of success in counterinsurgency.
16. Kahn 2002 argues for a similar conclusion, but based on a broader principle. He argues that in cases of highly asymmetric warfare – which he calls riskless warfare – it becomes morally

necessary to see these not properly as war, but rather as police actions. Thus, according to his argument, any use of AWS in asymmetric conflict would necessitate them being deployed for apprehension or disarmament efforts, rather than as killing machines.

17. Granted, in some battlespaces the likelihood of a wounded combatant receiving timely medical assistance can be low, meaning that otherwise non-lethal wounds may lead to death (see note 12 above). However, this only goes to show that even so-called "non-lethal" options still cause fatalities, and does not provide any positive reason for creating AWS which are designed to be lethal. In fact, the arguments in favor of AWS point to a further problem of designing them to make lethal targeting decisions; if they are better able to aim than humans and are also programed to aim for normally lethal parts of a human body, then it is likely that this would lead to field mortality rates far above what are currently the norm, which alone may render them in violation of international law (see, e.g. International Committee of the Red Cross 1997, esp. 22–24). Thanks to an anonymous reviewer for pressing me on this point.
18. Again, this is not to say the insurgent would with certainty be combat-ineffective, as there are a variety of things one could still do to contribute to the insurgency effort, but on average it is unlikely that an individual with a shattered knee (and presumably little to no access to proper medical facilities) will be capable of fighting counterinsurgent forces. Moreover, it is far from evident that the "added value" of killing that insurgent is morally weighty enough to merit making a lethal targeting decision in one instance, and even less so for making a general programing decision about an AWS' programing.
19. One might object that robots are not as good as LAWS proponents would have us believe, and might not be able to do all these things to such an advanced degree, thus undermining the preceding points. While this objection may be true, it argues against LAWS by denying the LAWS proponent one of their core premises. My purpose is to show that the argument for *lethal* AWS fails even when these premises are granted. Given that, I simply grant the proponent these strongest claims in favor of AWS.
20. Imagining such cases is a regular feature of the general philosophical literature on ethics and war. See, e.g. the discussions found in Nozick 1974, 34–35; Thomson 1991, 287; McMahan 2009, esp. ch. 4; or any of the thought-experiments from Frowe 2014, to name just a few.
21. For more on this point, see Section 6 below.
22. For example, St. Augustine said that "it ought to be *necessity*, and not your will, that destroys an enemy who is fighting you." (Augustine 2001, 217, emphasis added). See also Aquinas 2002, 241–242 and 218.
23. See preceding citations.
24. AP I, Art. 57.2.a is also considered to be a central statement of necessity, but it is concerned exclusively with minimizing/avoiding harm to civilians. For extensive treatment of the history of necessity in philosophy and law, see Ohlin and May 2016. See also Haque 2017, esp. ch. 7.
25. Terms like "superfluous injury" and "unnecessary suffering" do not by themselves provide clear guidelines for legal policy or military decision-making. In order to remedy this defect, the ICRC commissioned the SIrUS report, the aim of which was to provide clear medical demarcations of these terms. The report proposed four criteria for deciding whether a weapon would be deemed to cause "superfluous injury or unnecessary suffering", the second of which is relevant for the arguments here. Criterion 2 maintains that a weapon is to be considered to cause superfluous injury or unnecessary suffering if it causes "field mortality of more than 25% or a hospital mortality of more than 5%" (International Committee of the Red Cross 1997, 23). For the report's full recommendations, see esp. 22–28.
26. As stressed above, this is important to make clear as some discussions of AWS treat as interchangeable the rules governing flesh-and-blood troops and AWS (see, e.g. Anderson and Waxman 2012 and 2013). More conspicuously, some authors (most notably Roff 2013) base their arguments on the assumption that AWS will be "not simply weapons but a class of combatants" (Roff 2013, 352). However, AWS are not agents, and so cannot be thought of as combatants. At a basic level, AWS are no different than land-mines, given

that both require only human input in their deployment, after which they are capable of harming based on nothing more than their internal architecture, be it computer-based or mechanical.

27. These treaties represent the customary international law of necessity, and thus are binding upon all states and not merely those which are formally parties to them (Roberts and Guelff 2004, 54, 63). For the full statements of these treaties see Roberts and Guelff 2004, 53–57, 63–66.
28. The arguments of Mayer 2013 trade on this same legal point and provide support for the conclusions being argued for here.
29. For example, AI systems can better recognize and parse images than humans, and have been shown to be capable of operating at (at least) human levels in coordinated warfighting game scenarios. See Shoham et al. 2018.
30. By virtually all accounts there will be further constraints as well, such as that the violence serves some minimally moral or legal goal, that it does not violate other moral or legal principles, or that it is constrained by the principle of discrimination.
31. Thanks to anonymous reviewers for pressing me on all of these points.
32. If this assumption does not hold, then the objection fails, as it is then possible for an autonomous UAV to be programed to make non-lethal targeting decisions.
33. The AeroVironment WASP III is a small, light reconnaissance UAV currently in use by the U.S. Air Force and Marine Corps, as well as a number of other militaries.
34. See, e.g. Dinstein 2007; Sassòli and Olson 2008.
35. This, however, is at most a purely legal loosening of requirements, as there are compelling arguments that soldiers are morally required to opt for non-lethal weapons and tactics if these will be as effective. See, e.g. the philosophical texts cited in the Section 4 above, and especially the arguments of Mayer 2013.
36. See also the texts cited in note 10.
37. See, e.g. High Command Case No. 12, *United States vs. Leeb*, which maintained that military necessity does not allow one to do anything necessary to win a war (541).
38. This point is also strongly grounded in the previously cited Declarations (see note 27), as these treaties concern weapons which impose death or increased suffering upon soldiers who would otherwise be combat-ineffective but alive or less seriously harmed.
39. Some may object that this is itself problematic, as illegitimate regimes would then use robots to round up dissident or insurgent factions and then hold rigged trials and execute those combatants, with the apprehension and trial serving only to create an image of legitimacy where none is present. This is indeed a real fear worth bearing in mind, but it is an issue of political legitimacy and state oppression, not a question about the morality of robotics in warfare. There is no meaningful distinction between (a) an illegitimate state which uses superior technology and manpower to militarily strike insurgents and kill them, and (b) an illegitimate state which non-lethally apprehends insurgents, holds false trials where guilt is the only possible verdict, and then executes those insurgents. In either case illegitimate lethal force is used to protect a tyrannical regime, and opposition forces have little hope of opposing that force. If anything, the latter case (b) where robots are present is morally superior, as it is easier for non- or super-governmental organizations like Amnesty International and the U.N. to supervise trials than it is to supervise armed conflict, allowing for a higher probability of discovering illegitimate behaviors and abuses of power.

Acknowledgements

I must thank the Research Foundation – Flanders and Ghent University, whose support and assistance made this research possible. I would also like to thank Patrick Lin, who was kind enough to listen to some of my early ideas and point me toward potential objections and pitfalls, and Joke Meheus, who argued with me about the finer points of autonomy and harm. I am also greatly

indebted to Ryan Jenkins and Keith Abney, both of whom read an early draft and provided me with a wealth of insights. Finally, as always, I must thank my wife, who listened to the first "draft" of this work while we were on a bike ride, patiently enduring my long-winded lectures and surely missing out on the scenery as a result.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Research Foundation – Flanders (FWO).

Note on contributor

Nathan Wood is a PhD Fellow of the Research Foundation – Flanders (FWO), working from Ghent University in the fields of military ethics and formal ethics. His research focuses on the consequential just war constraints of *necessity* and *proportionality*, aiming to present a more coherent overall picture of those principles and to formulate them in formal terms. He has previously published in *Kriterion – Journal of Philosophy* and is a contributor to the upcoming collection *Just War Theory in an Age of Terror*.

References

Anderson, Kenneth, and Matthew Waxman. 2012. "Law and Ethics for Robot Soldiers." *Policy Review* 176. Accessed October 25, 2020. https://www.hoover.org/research/law-and-ethics-robot-soldiers.

Anderson, Kenneth, and Matthew Waxman. 2013. "*Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*." *Columbia Public Law Research Paper 13–35*. Accessed October 20, 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2250126.

Aquinas, Thomas. 2002. *Political Writings*. Edited and translated by R. W. Dyson. Cambridge: Cambridge University Press.

Arkin, Ronald C. 2009. *Governing Lethal Behavior in Autonomous Robots*. Boca Raton: CRC Press.

Arkin, Ronald C. 2010. "The Case for Ethical Autonomy in Unmanned Systems." *Journal of Military Ethics* 9 (4): 332–341.

Augustine. 2001. *Political Writings*. Edited by E. M. Atkins and R. J. Dodaro. Cambridge: Cambridge University Press.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Canning, John S. 2009. "You've Just Been Disarmed. Have a Nice Day!" *IEEE Technology and Society Magazine* 28 (1): 13–15.

Dinstein, Yoram. 2007. "The System of Status Groups in International Humanitarian Law." In *International Humanitarian Law Facing New Challenges*, edited by W. H. von Heinegg, and V. Epping, 145–156. Berlin: Springer.

Frowe, Helen. 2014. *Defensive Killing*. Oxford: Oxford University Press.

Green, Leslie C. 2017. *The Contemporary Law of Armed Conflict*. Manchester: Manchester University Press.

Guarini, Marcelo, and Paul Bello. 2012. "Robotic Warfare: Some Challenges in Moving from Noncivilian to Civilian Theaters." In *Robot Ethics: The Ethics and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. A. Bekey, 129–144. Cambridge: MIT Press.

Guetlein, Michael A. 2005. *Lethal Autonomous Weapons – Ethical and Doctrinal Implications*. Technical Report, United States Naval War College. Accessed October 25, 2020. https://apps.dtic.mil/dtic/tr/fulltext/u2/a464896.pdf.

Haque, Adil A. 2017. *Law and Morality at War*. Oxford: Oxford University Press.

Horton, Scott. 2006. "Kriegsraison or Military Necessity – the Bush Administration's Wilhelmine Attitude Towards the Conduct of War." *Fordham International Law Journal* 30: 576–598.

Hurka, Thomas. 2008. "Proportionality and Necessity." In *War: Essays in Political Philosophy*, edited by L. May, 127–144. Cambridge: Cambridge University Press.

International Committee of the Red Cross. 1997. *The SIrUS Project*. Technical Report, International Committee of the Red Cross. Accessed October 25, 2020. https://www.loc.gov/rr/frd/Military_Law/pdf/SIrUS-project.pdf.

Johnson, Aaron M., and Sidney Axinn. 2013. "The Morality of Autonomous Robots." *Journal of Military Ethics* 12 (2): 129–141.

Kahn, Paul W. 2002. "The Paradox of Riskless Warfare." *Philosophy and Public Policy Quarterly* 22 (3): 2–8.

Killmister, Suzy. 2008. "Remote Weaponry: The Ethical Implications." *Journal of Applied Philosophy* 25 (2): 121–133.

Kobayashi, Masahiko, and Paul F. Mellen. 2009. "Rubber Bullet Injury: Case Report with Autopsy Observation and Literature Review." *The American Journal of Forensic Medicine and Pathology* 30 (3): 262–267.

Lazar, Seth. 2012. "Necessity in Self-Defense and War." *Philosophy & Public Affairs* 40 (1): 3–44.

Lazar, Seth. 2014. "Necessity and Non-Combatant Immunity." *Review of International Studies* 40 (1): 53–76.

Lin, Patrick, George A. Bekey, and Keith Abney. 2008. *Autonomous Military Robotics: Risk, Ethics, and Design*. Technical Report, California Polytechnic State University, prepared for the Department of the Navy. Accessed October 26, 2020. http://ethics.calpoly.edu/ONR_report.pdf.

Lokhorst, Gert-Jan, and Jeroen van den Hoven. 2012. "Responsibility for Military Robots." In *Robot Ethics: The Ethics and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. A. Bekey, 145–156. Cambridge: MIT Press.

Mayer, Chris. 2013. "Minimizing Harm to Combatants: Nonlethal Weapons, Combatants' Rights, and State Responsibility." In *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century*, edited by F. Allhoff, N. G. Evans, and A. Henschke, 301–311. New York: Routledge.

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

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2 (2): 175–220.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell.

Ohlin, Jens D., and Larry May. 2016. *Necessity in International Law*. Oxford: Oxford University Press.

Paust, Jordan J. 1974. "Weapons Regulation, Military Necessity and Legal Standards: Are Contemporary Department of Defense Practices Inconsistent with Legal Norms?" *Denver Journal of International Law and Policy* 4: 229–235.

Purves, Duncan, Ryan Jenkins, and Bradley J. Strawser. 2015. "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons." *Ethical Theory and Moral Practice* 18 (4): 851–872.

Roberts, Adam, and Richard Guelff, eds. 2004. *Documents on the Laws of War*. 3rd ed. Oxford: Oxford University Press.

Robillard, Michael. 2017. "No Such Thing as Killer Robots." *Journal of Applied Philosophy* 35 (4): 705–717.

Roff, Heather M. 2013. "Killing in War: Responsibility, Liability, and Lethal Autonomous Robots." In *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century*, edited by F. Allhoff, N. G. Evans, and A. Henschke, 352–364. New York: Routledge.

Roff, Heather M. 2014. "The Strategic Robot Problem: Lethal Autonomous Weapons in War." *Journal of Military Ethics* 13 (3): 211–227.

Roff, Heather M. forthcoming. "An Ontology of Autonomy and Autonomous Weapons Systems." In *The Ethics of Autonomous Weapons*, edited by C. Finkelstein, D. MacIntosh, and J. D. Ohlin. Oxford: Oxford University Press.

Roff, Heather M., and David Danks. 2018. "'Trust but Verify': The Difficulty of Trusting Autonomous Weapons Systems." *Journal of Military Ethics* 17 (1): 2–20.

Sassòli, Marco, and Laura M. Olson. 2008. "The Relationship between International Humanitarian and Human Rights Law Where It Matters: Admissible Killing and Internment of Fighters in Non-International Armed Conflicts." *International Review of the Red Cross* 90 (871): 599–627.

Sharkey, Noel. 2010. "Saying 'No!' to Lethal Autonomous Targeting." *Journal of Military Ethics* 9 (4): 369–383.

Shoham, Yoav, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz, and Zoe Bauer. 2018. *The AI Index 2018 Annual Report*. Technical Report. Accessed October 25, 2020. http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf.

Sparrow, Robert. 2007. "Killer Robots." *Journal of Applied Philosophy* 24 (1): 62–77.

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

Thomson, Judith J. 1991. "Self-defense." *Philosophy & Public Affairs* 20 (4): 283–310.

U.S. Army. 1956. *FM 27-10 The Law of Land Warfare*. Technical Report, United States Army. Accessed October 26, 2020. https://www.loc.gov/rr/frd/Military_Law/pdf/law_warfare-1956.pdf.

U.S. Army/U.S. Marine Corps. 2014. *FM 3-24 Insurgencies and Countering Insurgencies*. Technical Report, United States Army. Accessed October 26, 2020. https://fas.org/irp/doddir/army/fm3-24.pdf.

U.S. Department of Defense. 2016. *Fiscal Year 2017 Budget Request*. United States Department of Defense. Accessed October 25, 2020. https://www.defense.gov/Newsroom/Releases/Release/Article/652687/department-of-defense-dod-releases-fiscal-year-2017-presidents-budget-proposal/.

U.S. Department of Defense. 2017. *DoD Directive 3000.09*. United States Department of Defense. Accessed October 25, 2020. https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

U.S. Department of Defense. 2018. *DoD Dictionary of Military and Associated Terms*. Technical Report, United States Department of Defense. Accessed October 25, 2020. https://www.hsdl.org/?view&did=820026.

U.S. Marine Corps Training Command. 2015. *Law of War/Introduction to Rules of Engagement*. Technical Report, United States Marine Corps. Accessed October 20, 2020. https://www.trngcmd.marines.mil/Portals/207/Docs/TBS/B130936%20Law%20of%20War%20and%20Rules%20Of%20Engagement.pdf.

U.S. Surgeon General. 2006. *Mental Health Advisory Team (MHAT) IV: Operation Iraqi Freedom 05-07, Final Report*. Technical Report, Office of the Surgeon General, United States Army Medical Command. Accessed October 20, 2020. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB2010103335.xhtml.

Wood, Nathan G. forthcoming. "Necessity and the 'War on Terror': Lessons from the U.S. Army/Marine Corps Counterinsurgency Manual." In *Just War Theory in an Age of Terror*, edited by Ty Hawkins and Andrew Kim. [Publisher to be decided.]