Abstract

The thesis offers a comprehensive argument in favor of a regulationist approach to autonomous weapon systems (AWS). AWS, defined as all military robots capable of selecting or engaging targets without direct human involvement, are an emerging and potentially deeply transformative military technology subject to very substantial ethical controversy. AWS have both their enthusiasts and their detractors, prominently advocating for a global preemptive ban on AWS development and use. Rejecting both positions, the author outlines a middle-of-the-road regulationist approach that is neither overly restrictive nor overly permissive. The disqualifying flaws of the rival prohibitionist approach are demonstrated in the process.

After defining the core term of autonomy in weapon systems, the practical difficulties involved in applying an arms control regime to AWS are analyzed. The analysis shows that AWS are an extremely regulation-resistant technology. This feature when combined with their assumed high military utility makes a ban framework extremely costly to impose and enforce. As such it is ultimately very likely to fail at the benefit of the most unscrupulous international actors and at a very substantial risk to those abiding with international law. Consequently, to be ethically viable, a prohibitionist framework would need to offer substantial moral benefits impossible to attain through the rival regulationist approach. The remainder of the thesis undertakes to demonstrate that this is not the case.

Comparing the considerations of military and strategic necessity to humanitarian concerns most commonly voiced by prohibitionists requires finding a common denominator for all values being referred to. Consequently, the thesis proceeds to show that both kinds of concerns are ultimately reducible to respect for basic human rights of all stakeholders, and so that the prohibitionist and regulationist approach may ultimately be compared in terms of consequences their adoption would generate for basic human rights realization.

The author then evaluates both the potential humanitarian benefits, and the potential humanitarian hazards of AWS introduction. The benefits of leaving frontline combat to machines are outlined, with the unique kinds of suffering that would be abolished by such a development being described in detail. The arguments against AWS adoption are then divided into three classes: arguments related to alleged impossibility of compliance with The Laws of Armed Conflict, non-consequentialist and broad consequentialist arguments.

This analysis, which comprises the greater part of the entire thesis, shows that the concerns behind compliance arguments are indeed substantial and have to be accommodated via a complex framework of best practices, regulations and localized restrictions on some kinds of AWS or AWS use in particular environments. They do not, however, justify a universal ban on using all the diverse forms of AWS in all environments. Non-consequentialist objections are found either reducible to other classes of arguments or thoroughly unconvincing, sometimes to the point of being actually vacuous. Broad consequentialist concerns are likewise found to be accommodable by regulation, empirically unfounded or causally disconnected from the actions of legitimate actors acquiring AWS, and therefore irrelevant to the moral permissibility of such actions.

The author concludes that the proponents of prohibitionism are unable to point to moral benefits substantial enough to justify the costs and risks inherent in the approach. A global ban is, in fact, likely to have a worse humanitarian impact that well-regulated AWS adoption even if the strategic risks are disregarded. On the other hand, the analysis shows that there indeed exists an urgent need to regulate AWS through a variety of technological, procedural and legal solutions. These include, but are not limited to, a temporary moratorium on anti-personnel AWS use, development of internationally verified compliance software and eventual legal requirement of its employment, a

temporary moratorium on AWS proliferation to state actors and a ban on their proliferation to non-state agents.

Table of Contents

Introduction

What counts, and what does not count as a technological revolution in military affairs is a subject of controversy, and for good reasons. Yet very few would deny that such a revolution happened at least twice within the last hundred years, and that at least these two times the impact was ethically significant.

First came the warplane, humanity's perennial dream come true. While other technological novelties – the machine gun, long range artillery, poison gas – combined to create the hell of trench warfare, knights of the skies soared above it all to engage in chivalrous duels. The warplane seemed innocent, and not terribly significant. The War to End All Wars gave way to a dozen smaller ones, aviation's heroics continued to fascinate the public.

Then smart men realized the bomber will always get through. They even had some time to dream of precision bombing, forty years too early (Gladwell 2021).

And then the cities burned. There were nights when tens of thousands of civilians burned alive within a few hours. Horrors that could not and would not be effected if the men effecting them could see them from up close, and that sickened them to the bone even when witnessed from thousands of feet.

We agreed we would not do it again, not with normal bombs anyway. Hanoi and Haiphong were targeted for their war industry, not for their people, though the bombing was ineffective and gruesome by modern standards. The relationship between the two outcomes led to the development of precision guided munitions, and then to drones. The dream of the Bomber Mafia has been realized, and today when an airstrike ends up in disproportional slaughter, it is localized malice or criminal negligence, not national policy or technological necessity that is to blame. The warplane, we may say, finally turned out ok. If the technology was to disappear from the face of the Earth, the world would be a worse place, ethically speaking.

I have a firm belief this state of affairs could have been achieved with many fewer mothers smothering their children so they would not burn alive.

I have the same belief about the nuclear revolution. Whether one is contended with the present reality of nuclear balance or not, I think most would agree we could have arrived at the same result by a safer road (Kaplan 2020). My point is that the way in which humanity handles a revolution in military affairs matters very much, militarily, politically, economically, socially and, most of all,

ethically. I am making this point because whether or not we welcome the opportunity, we get to try again.

Autonomous weapon systems – artificially intelligent robots capable of finding and attacking targets on their own – were an imminent technological possibility when the idea of writing a doctoral thesis on the ethical issues surrounding them first occurred to me in 2016. As of December 2021, platforms that can be uncontroversially termed AWS most probably already exist, and have for some time[1]. They have been made possible by steady progress in computer science, robotics and related fields and follow in the wake of the already highly transformative, though not yet fully implemented, drone revolution. The new revolution in military affairs promises (or threatens) to transform warfare in a way comparable with the invention of the warplane or nuclear weapons. The resulting repercussions for the political, strategic and humanitarian landscape are likely to be comparable as well, especially once the technology fully develops and when measured over comparable timescales[2].

It stands to reason we should evaluate both the threat and the promise, the ways in which they are tied and the ways in which they are independent, and minimize the former while maximizing the latter. Our reaction should be informed not only by a clear and honest vision of the best possible outcome, but also by the likelihood of achieving this outcome, and alternative tolerable outcomes, in the world as it is today. This work is an attempt to compare two alternative broad courses of action from an ethical perspective, as opposed to legal, pragmatic, or administrative. I will try to show why one of them should be preferred as better in the ethical sense, and so enacted or at the very least left open as a possibility. This is not to say that legal, pragmatic, or administrative perspectives are ethically irrelevant. They are. However, they are not self-standing. They are aspects of an ethical perspective, and so they must to be discussed in a work whose ambition is ethical. Along the way I will sketch, in broad strokes, though still more specifically than others before me, how this might be done.

What are these two courses of action? One is to pursue a very ambitious and most probably unprecedented agenda of scientific research, policy deliberation, arms control and military reform. This agenda would be intended to ensure, to the greatest extent possible, that legitimate political actors, rather than choosing between national security and humanitarian concerns amidst the chaos of a runaway arms race, master and utilize this new technology to arrive at the equilibrium that features both enhanced security and an improved humanitarian landscape. This sounds very ambitious and idealistic, but many argue we have to try. I call various approaches that meet this rough description regulationist ones.

---

1 What was likely the first instance of their combat use reportedly took place in April 2020 (UNSC Panel of Experts on Libya 2021, 17)

2 If AWS never come to exist because of an effective ban, the impact of this outcome will of course be equally significant and long lasting, although different.

The other frequently advocated course of action is to embark on an even more ambitious and most definitely unprecedented agenda to nip the coming revolution in military affairs in the bud by enacting and enforcing a global preventive ban on AWS. I call this approach prohibitionism[3].

There is a third possible approach – that of ignoring the coming revolution in military affairs and not trying to shape it in any way. This approach might be called permissivist, as not regulating an emerging technology is equivalent to permitting its use in any way possible. Thus, we might see the normative options regarding the use of AWS as spanning from permissivism to prohibitionism with various versions of regulationism in between.

As evidenced by the course of the two previous military revolutions, permissivism is not an ethically acceptable approach. Nor is it a popular one. Paul Scharre, who has personally interviewed most key voices in the AWS debate has stated that "in the nearly ten years I have spent working on the issue of autonomous weapons, almost every person I spoken with has argued there ought to be some limits on what actions machines can take in war, although they draw this line in very different places" (2018, 346). On the basis of my own research, I concur[4]. Permissivism is the option most unlikely to be chosen in the ethical debate on the use of AWS. The real options are a regulationist or prohibitionist approach.

Is it the regulationist or prohibitionist approach then? When I embarked on writing this thesis, my goal was to make the strongest possible case for the regulationist approach. In researching and writing it, I became acutely aware of the difficulties inherent in crafting and executing this approach. While initially I was positive the AWS revolution could be made to benefit humanity, having spent five years pondering this issue I feel that some of my sunny optimism has evaporated, and my views shifted towards advocating for harsher restrictions and more robust arms control efforts.

Yet I have also understood how practically infeasible, strategically risky and ethically unfounded the prohibitionist approach was. Even more importantly, I understood the AWS debate was in large part ill-focused and ill-ordered, incorrectly assigning the burden of proof and largely failing to account for practically and ethically crucial context.

In the remainder of this introduction, I will broadly outline my case for this being true, simultaneously outlining my plan for this thesis. I will then situate my work among other book-size treatments of the topic and discuss my criteria for literature inclusion. This will be followed by review of some ethical concerns and an important disclaimer as to this work's character; finally, I will briefly

3 While regulationism admits of many versions, prohibitionism has essentially only one version, as prohibitionism admitting of significant exceptions becomes a version of regulationism.

4 It is important to stress that there are, in fact, serious and numerous voices stating that no new international law has to be created in order to accommodate AWS emergence. Yet this is not tantamount to the do nothing approach – all these voices advocate robust scientific, organizational and national-level regulatory efforts.

argue that my thesis would still contribute philosophical and scholarly value even if my argument was proven wrong.

As the aim of my work is determining which of the two policy options is ethically better, it makes sense to start my assessment by asking what would it take to implement each approach, and whether these approaches are even possible to implement. After all, there is no merit in deliberating on options that are not really available, and the cost of implementing policy options should definitely figure in their moral assessment, especially if part of this cost consists of taking moral risks or sacrificing human lives – as is sometimes the case with arms control policies. This is why, right after settling on a definition of AWS and explaining this choice in Chapter I, I look at the feasibility and cost of implementing regulationist and prohibitionist approaches in Chapter II. Both are, by historical standards, extremely ambitious and likely to partially fail in their goals. As the more restrictive one, the prohibitionist approach may be suspected of being more costly (in a financial, political and humanitarian way) and less likely to succeed. I argue this is indeed the case, partially basing my assessment on the assumption of AWS' decisive military potential.

I am not a technologist or an expert in military science, and therefore I will not attempt to argue at length that AWS would be a militarily transformative technology once mature enough, nor will I try to ascertain the exact extent of their military utility. It is not my place, and not in my competences, to try to settle this. Until demonstration on the battlefield or at least proving ground conditions, this matter cannot be said to have been definitively proven anyway. Yet I think that if there was even a 10% chance that AWS will be to XXI century warfare what the warplane was to the XX century, making such an assumption and seeing where it leads one would be a worthwhile exercise. As will become evident in subsequent chapters, chances that these machines able to rid warfare of human limitations will prove militarily decisive are in fact much higher – in fact, they are probably very close to 100%. Once human flight was proved to be possible, it was not very plausible to assume planes will never surpass birds, and that they will not surpass birds by far. The same is most probably true of AWS and human combatants. Still, it is not my role to immerse myself in this issue, which is not an ethical problem. Let me just clarify my assumption in no uncertain terms.

I assume AWS will offer military advantage over manned or remotely controlled platforms substantial enough that to eschew their use would mean losing to a peer competitor in any military confrontation, just like eschewing the use of aircraft or the use of precision guided munitions would mean losing to a peer competitor today[5].

5 Note that "peer-competitor" does not denote a hypothetical opponent who is precisely evenly matched (in which case foregoing even the smallest advantage would have this effect), but an opponent who is assumed to be on roughly the same level of military capacity.

Now, if AWS indeed grant their users decisive military potential, then they cannot be treated like landmines or blinding laser have been treated, in the sense that unilaterally renouncing them is not an option. This, as I will argue, radically impacts both the likelihood of success in banning the technology, but also the risks involved in a possible (indeed, likely) collapse of the ban framework. The actors who choose to comply with the ban risk being militarily dominated by a successful violator, especially if such a violation may be kept secret up to the point when the AWS in question actually become operational. As I will also argue, other features of AWS technology make such secretive violation, or sneak out, highly likely. Simultaneously yet other features make AWS particularly attractive to actors generally uncommitted to values behind arms control, especially authoritarian states and state sponsors of terrorism. All in all, the likely collapse of the ban framework carries a substantial risk of either the lasting military domination of such amoral actors or a runaway arms race in which morally legitimate actors are playing catch up, with predictable consequences for the world peace or efforts to make AWS LOAC-compliant.

This assessment of the additional risks of the prohibitionist approach allows me to establish a standard to be met by the prohibitionists and regulationists alike. Given that the prohibitionist approach carries with it additional and substantial risks, its advocates should prove that these risks are indeed worth taking. In other words, AWS prohibition, as opposed to their regulation, would have to bring benefits (or avoid harms) large enough to at least equal the harms inherent in these risks. This not only helps to establish a burden of proof in the debate, but grounds my ethical inquiry in the context of the military, strategic and geopolitical realities of the subject, in which actual policymakers – the ultimate addressees of ethical arguments – operate.

Are the risks involved in attempting AWS prohibition comparable to the postulated harms of allowing (some of) them to be used, though? In order to be able to cash all these arguments out and compare them to each other, we need a common currency. In Chapter 3 I argue that such a common currency of military and political ethics in general, and of the AWS debate in particular, can be found in the concept of basic human rights. Contemporary just war theory, the theoretical basis of military ethics, has established that military (or geopolitical) necessity and humanitarian concerns do not involve a clash of incommensurable, and so tragically opposed values, but are both expressible in the language of basic human rights that both a legitimate state actor and a humanitarian advocate attempt to respect and protect, or realize.

Actors legitimately wielding military power derive their legitimacy from their status of institutions necessary for basic human rights realization; sovereignty and its prerogatives are granted at the service of such rights realization, and attrited or lost to when a state reneges on this function. Just war theory in its current form is in fact nothing but an account of the conditions under which it is permissible for legitimate states to use organized violence to protect basic human rights.

Consequently, military necessity, when genuinely existing, can be cashed out in the same currency as humanitarian concerns; to say that humanitarian concerns trump military necessity in a specific case is to say that that the rights of more persons, or more important rights would be violated by a certain military action than would be preserved by it.

The work conducted in Chapter 3 will thus allow us to see that the national security risks indicated in Chapter 2, and various alleged humanitarian threats caused by AWS adoption discussed in Chapters 5-9 do have a common denominator, as do potential humanitarian benefits of AWS use, discussed in Chapter 4. It will also serve to explain and undergird the specific regulations that I do postulate bind AWS development and use. While in various chapters I will be using seemingly incommensurable vocabularies of LOAC, just war theory, strategic discourse etc., all these will ultimately be expressible in the common axiological currency of basic human rights.

The rest of the thesis will proceed as follows: in Chapter 4, as mentioned, I will discuss potential humanitarian benefits of AWS use, arguing that widespread yet well-regulated AWS use could not only make certain kinds of harm inherent in war less common, but also that it could in fact eliminate certain kinds of intense suffering unique to the experience of fighting in war. In Chapters 5-7 I will wrestle with the extremely complex issue of the possibility of LOAC-compliant AWS use. Chapter 5 will ground the issue in the necessary context of technology and military practice; Chapter 6 will deal with the issue of compliance with the Principle of Distinction, while Chapter 7 will discuss the possible ways of complying with other LOAC/Ius in Bello principles. Chapter 8 will treat on non-consequentialist arguments, that is arguments purporting to show that AWS use is inherently wrong regardless of the particular strategy of use or the technical characteristics of a specific platform – and regardless of the user's ability to comply with LOAC. Finally Chapter 9 will analyze the broad consequentialist arguments against AWS adoption, that is arguments from long term, possibly indirect consequences of their introduction.

While this outline may suggest the present work will focus on making a negative case against the ban and advocate for a regulationist approach only indirectly by showing the alternative to be infeasible, such a conclusion is only partially true. Through the process of weighing various moral charges levied against AWS I will be gradually arriving at the conception of ethically justified regulatory framework, while also creating a model of not only morally permissive, but morally praiseworthy use of some kinds of AWS, boosting the positive case for the technology's introduction.