Georgia State University

ScholarWorks @ Georgia State University

Philosophy Theses   Department of Philosophy

Spring 5-10-2019

Arguments for Banning Autonomous Weapon Systems: A Critique

Hunter B. Cantrell
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

ARGUMENTS FOR BANNING AUTONOMOUS WEAPON SYSTEMS: A

CRITIQUE.

by

HUNTER CANTRELL

Under the Direction of Andrew Altman, PhD

ABSTRACT

Autonomous Weapon Systems (AWS) are the next logical advancement for military technology. There is a significant concern though that by allowing such systems on the battlefield, we are collectively abdicating our moral responsibility. In this thesis, I will examine two arguments that advocate for a total ban on the use of AWS. I call these arguments the "Responsibility" and the "Agency" arguments. After presenting these arguments, I provide my own objections and demonstrate why these arguments fail to convince. I then provide an argument as to why the use of AWS is a rational choice in the evolution of warfare. I conclude my thesis by providing a framework upon which future international regulations regarding AWS could be built.

ARGUMENTS FOR BANNING AUTONOMOUS WEAPON SYSTEMS: A CRITIQUE

by

HUNTER CANTRELL

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2019

ARGUMENTS FOR BANNING AUTONOMOUS WEAPON SYSTEMS: A CRITIQUE

by

HUNTER CANTRELL

Committee Chair: Andrew Altman

Committee: Andrew I Cohen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

DEDICATION

This thesis is dedicated to my loving wife Tiffany and my two wonderful boys Holden and Bryson. Without you, I would be nothing.

ACKNOWLEDGMENTS

I would like to acknowledge the faculty and staff of the Philosophy Department, especially Dr. Andrew Altman and Dr. Andrew I Cohen: without your dedication, this thesis would not have been successful. I would also like to thank all the professors with whom I have had the privilege of taking a course: thank you for putting up with me. I would also like to thank those in the US Army who saw a spark in me and thought I was worthy of the task. Thank you all.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

- AWS- Autonomous Weapon System(s)

- ICC- International Criminal Court

- IHL- International Humanitarian Law

- LoW- Law of War

- RoE- Rules of Engagement

1   INTRODUCTION

The anticipated–imminent"[1]–rise of fully autonomous weapons systems has led to a call for a "campaign to stop killer robots."[2] This call has so far been backed by twenty-three nations[3] and has grown in perceived legitimacy since the campaign's inception in 2013.[4] These efforts are driven in part by a fear of a future in which all (or at least the lethal) functions of human soldiers have been removed from the battlefield and replaced with a lethal form of artificial intelligence. However, there is no reason to think that the fear is a realistic one, contrary to the Hollywood depiction of apocalyptic Terminators roaming city streets.

Nonetheless, there are reasoned arguments against the development and implementation of fully autonomous weapons systems. We can classify the arguments against such systems into two main categories: an argument from "responsibility" and an argument from "agency." Roughly, the argument from responsibility is concerned with the chain of responsibility for the use of autonomous weapons systems (AWS) and who ought to bear responsibility should something go wrong. And the argument from agency questions why we should exchange human agency in a lethal action for an artificial agency and the implication this exchange has for human dignity and the possibility of mercy.

In this thesis, I summarize and evaluate these arguments in turn and show that both arguments are insufficient for an outright ban on AWS. I contend that the arguments are sufficient to support a framework, built on existing international humanitarian law, to *regulate but not ban* the

---

[1] There is a considerable lack of consensus on what counts as imminent: 5, 10, 50, 100 years? This lack of clarity is rampant throughout the field of emergent autonomous weapons systems.
[2] Campaign to Stop Killer Robots, "About Us."
[3] It should be noted that all of the nations that have so far joined the campaign to "stop killer robots" are unlikely to be the nations that have the technological capability, infrastructure, or military "necessity" to build such devices. The list includes: Algeria, Argentina, Bolivia, Brazil, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, Guatemala, Holy See, Iraq, Mexico, Nicaragua, Pakistan, Panama, Peru, State of Palestine, Uganda, Venezuela, Zimbabwe (Campaign to Stop Killer Robots).
[4] The campaign has met every year since 2013, hosted by the United Nations Office in Geneva, Switzerland.

development and employment of AWS. I then argue that the use of AWS is a rational choice for military and political leaders. I also outline a reasonable framework based on the current Geneva Conventions and international humanitarian law for regulating AWS.

1.1 Why is it essential to discuss Autonomous Weapon Systems?

Autonomous Weapons Systems are the—"next frontier"—of modern military technology. As the speed of decision making becomes faster due to increasing computerization, the ability of the human mind to keep up is lagging. The OODA loop was first coined by USAF Pilot and air power strategist Colonel John Boyd.[5] It stands for Observe, Orient, Decide, Act. The basic premise is that if a fighter pilot can get inside the OODA loop of an enemy fighter pilot, he could defeat him. The OODA loop then is a short decision matrix process in which a person (or in the future a machine) can decide on a certain action. It is likely that these OODA loops will continue to narrow as technology increases. The narrowed OODA loop of the future lends itself naturally to artificial intelligence that can process inputs and compute multiple decisions in real time much faster than any human. The drive to make—"smarter" and faster—weapons can clearly be seen in the evolution of fighter jets (USAF F-35 II and the Chinese Chengdu J-20 are but two examples),—"smart bombs"—(precision-guided munitions such as the Excalibur[6]), and the increased utilization of computer-assisted decision making tools in military planning.[7]

In response to the increased complexity of modern warfare and the decrease in the available population deemed fit to serve[8], increased automation in current weapons systems and a drive to implement fully autonomous weapons systems have been the remedy. Current systems such as the

---

[5] Boyd 1987, 383.
[6] US Army Acquisition Support Center, "Excalibur Precision 155MM Projectiles."
[7] This should not be construed as currently using computers *to* make command decisions, but rather the reliance of computer-based systems to provide more, better, and/or clearer information to a commander. These systems included everything from world-wide cargo tracking, enemy and friendly unit battlefield tracking, electronic systems monitoring, etc. As the civilian world has become more digitally reliant, so to has the military.
[8] See the Military Leadership Diversity Commission Issue Paper #2 "Outreach & Recruiting."

PHALANX (deployed on US Aegis Cruisers) allows a semi-autonomous system to perform critical ship protection functions, freeing up the limited crew to perform other essential duties[9]. The increasing use of such technologies is the first stepping stone in the implementation of a system that will have the ability to select "targets and [deliver] force without any human input or interaction."[10] The implementation of AWS is potentially worrying on many fronts, as alluded to earlier, but this development has the beneficial potential of reducing the unforeseen or collateral damage that accompanies warfare as currently practiced.

1.2   What are Autonomous Weapon Systems?

Before delving into the debate regarding the implementation of AWS, it is essential to understand just what an autonomous weapon is. The academic and diplomatic communities, unsurprisingly, have yet to settle upon a commonly agreed upon working definition of AWS. Currently accepted diplomatic definitions vary widely: some include weapons that today are considered automated, though not autonomous, and others include only systems wholly separated from human interaction. There are also definitions that require specific functionalities or technologies to be present. One example is from the Government of the Netherlands which defines an AWS as "a weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention."[11] This definition could easily encompass weapons that today are not considered to be autonomous, such as a cruise missile[12]. A cruise missile is designed to seek a specific programmed target (a heat signature, laser

---

[9] Raytheon, "Phalanx Close-In Weapon System."
[10] ICRC 2015, 6.
[11] United Nations Institute for Disarmament Research [UNIDIR] 2017, 23
[12] This example could also extend to any other "precision guided munition." It should be noted that current law according the Geneva Convention mandates the use non-indiscriminate weapons which can be generally interpreted as meaning it endorses using precision munitions over indiscriminate ones and prohibits "[e]mploying weapons, projectiles and material and methods of warfare … which are inherently indiscriminate in violation of the

designation from a soldier on the ground, or some other predesignated marker) and once a human has given the order to deploy the system, it cannot be recalled.

The Government of France approaches the definition of an AWS in terms of "what it is not"[13] stipulating:

> Lethal autonomous weapons are fully autonomous systems … LAWS should be understood as implying a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command. … The delivery platform of a LAWS would be capable of moving, adapting to its land, marine, or aerial environments and targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation. … LAWS would most likely possess self-learning capabilities.[14]

The French definition differs from the Dutch, in that it requires that an AWS (or LAWS) have no connection with a military (read human) chain of command, including in mission planning and final issuance of orders to use the AWS. This definition would severely limit what could be called an AWS as it is likely that there will be some meaningful human interaction somewhere in the "kill chain." The French definition then is so narrow as to make it practically untenable as a foundation upon which to build practical diplomatic and academic discussions.[15]

While adopting elements of the functionalist approaches used by the French, the United States' definition is distinct because it distinguishes autonomous from semi-autonomous systems. The United States (through the Department of Defense) defines an AWS as,

> A weapon system, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system but can select and engage targets without further human input after activation.[16]

---

international law of armed conflict, provided that such weapons, projectiles and material and methods of warfare are the subject of a comprehensive prohibition and are included in an annex to this Statute" (International Criminal Court Article 8(2)(b)(xx) 1998).

[13] United Nations Institute for Disarmament Research [UNIDIR] 2017, 24

[14] Ibid, 24.

[15] This can also be said of the current Chinese designation of a AWS having a level of autonomy which lacks "human intervention and control during the entire process of executing a task" (Government of China 2018).

[16] " United Nations Institute for Disarmament Research [UNIDIR] 2017, 30.

The US definition then is more practically useful than the French or Dutch for it gives a detailed account of what is an AWS is and, importantly, excludes by implication systems that should not count as an AWS.[17] The definition does not include systems that are in place today that would be considered automated or semi-autonomous[18] (PHALANX, C-RAM[19], etc.) nor does it include "fire and forget" systems, which could include the aforementioned cruise missile. It is also important to note that the US definition allows an AWS to act without the intervention of a human operator but does *not* preclude interaction from a human operator. This is important as it falls clearly in line with the US DoD Policy Directive 3000.09, that states that "Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[20] This policy directive hints at the collaborative nature that the US envisions for the implementation of artificial intelligence and AWS. The US policy definition also "places the focus of what constitutes autonomy at the level of *decision* rather than on the presence or absence of a particular technology."[21]

1.3   What does current International Humanitarian Law say regarding AWS?

Modern military conflict is governed by a set of norms and principles generally agreed upon as the basis for conducting a war. These norms and principles are built upon the long history of

---

[17] As mentioned, the US provides a definition of what it defines as *not* an AWS, "A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator. This includes: (a) semi-autonomous weapon systems that employ autonomy for engagement related functions, including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement. (b) "Fire and forget" or lock-on-after launch homing munitions that rely on TTP (tactics, techniques and procedures) to maximize the probability that only the targets within the seeker's acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator" (UNIDIR 2017, 31).

[18] It should be noted though, that systems such as the PHALANX and C-RAM can be placed into a setting which is analogous to fully autonomous.

[19] Program Executive Office- Missiles & Space, "Counter-Rocket, Artillery, Mortar (C-RAM)."

[20] Department of Defense, Directive 3000.09.

[21] UNIDIR 2017 31.

the just war theory. Prior to the twentieth century, these norms and principles were not codified but rather a set of agreed upon customs and practices.[22] For instance, it was generally considered the "gentlemanly" practice to rescue as many enemy sailors as possible once a naval battle had concluded. This practice was not required by treaty or other formal international agreement but rather was an accepted practice that evolved over centuries of naval warfare. Though it had been common practice, with the proliferation of submarine warfare, the practice was abandoned. After World War II, during the Nuremberg trials, the Allies accused Admiral Donitz of committing crimes against humanity for the Nazis tactic of sinking enemy vessels and refusing to surface to rescue the survivors. The prosecution failed though, as it was determined that through the evolution of naval warfare it had become common custom for submarines not to pick up surviving, stranded sailors due to extreme operational risk.

International Humanitarian Law (IHL) in the form of the Geneva Conventions and its Additional Protocols does not explicitly mention AWS. What can be useful for developing rules regulating AWS, though, there are IHL prohibitions on certain sorts of weaponry. The Convention on Certain Conventional Weapons (CCW) bans the use of certain inherently indiscriminate forms of weapons. This prohibition implies a preferential status of—precision—or "smart" weapons over indiscriminate or "dumb" ones. To follow the CCW then, an AWS would have to be developed and utilized in a fashion in which its use would not be "inherently indiscriminate." What then is a weapon or weapon system that is inherently indiscriminate? A prime example would be a "dumb" anti-personnel landmine or a poison gas. Both types of weapons are indiscriminate because upon

---

[22] While it is true that some elements of international law are still predicated on customs and norms, there are now treaty-based rules in place, such as those specified in the Geneva Conventions (1949) and in the Statute of the International Criminal Court. Prior to the establishment of such bodies as the ICC, most nations only had the assurances of warfare or mutually supportive defensive alliances to enforce compliance with agreed upon customs and norms, and some nations, including the US are not state parties to the ICC.

deployment they cannot correctly distinguish between a legitimate and a non-legitimate military target. The land mine will kill or maim any who are unfortunate to step on it, even decades later. Developing an AWS that can accurately discriminate is a difficult task, but not one that is insurmountable.

2   OBJECTIONS TO AWS: THE ARGUMENTS FOR A BAN

Since the invention of the crossbow in the 13th century[23], there have been calls for the limitation of acceptable weaponry in "civilized warfare." Technological developments have pushed nations to seek the latest cutting-edge weaponry to gain even the slightest advantage over their enemy. These developments have included the machine-gun, the landmine, and poison gases. Some of the developments, such as machine-guns, have stayed within the bounds of acceptable military hardware. Others such as the landmine have been heavily regulated, and their use is internationally discouraged.[24] The third category of weapons, including poison gases, has been expressly forbidden under international treaty.

Since 2013, the Campaign to Stop Killer Robots has advocated for a total ban on the development and employment of lethal autonomous weapon systems.[25] The group has advocated at the state and international level for a coalition of UN member states, non-governmental organizations, businesses, and private citizens with the express purpose of banning lethal AWS. The Campaign, like many persons in academia worries AWS, will cross a "moral threshold,"

---

[23] From the Papal Canon issued after the Second Council of the Lateran "We prohibit under anathema that murderous art of crossbowmen and archers, which is hateful to God, to be employed against Christians and Catholics from now on." (Papal Encyclicals Online, "The Second Council of the Lateran")

[24] Anti-personnel landmines are explicitly prohibited under the Quebec treaty:… anti-tank and anti-ship mines are still acceptable. Mines that have a short life span (such as becoming inert after X hours of deployment) are acceptable.

[25] "The Campaign to Stop Killer Robots calls for a pre-emptive and comprehensive ban on the development, production, and use of fully autonomous weapons, also known as lethal autonomous weapons systems or killer robots" (Campaign to Stop Killer Robots Call to Action, https://www.stopkillerrobots.org/call-to-action/).

sacrificing the dignity of human life for the efficiency of automated warfare. The arguments for a ban on "killer robots" can roughly be grouped into two categories, the "Responsibility" Argument, and the "Agency" Argument.

2.1   The "Responsibility" Argument: Who can be held responsible?

One of the great worries of those against the use of AWS is that we may create robots that can target and kill humans without being able to "justly" hold a moral agent responsible. Some have called the problem of justly assigning responsibility for an action taken by an AWS an "accountability gap."[26] To be sure, if we can trace an action of an AWS directly to the person who authorized the mission (for instance, a rogue commander who programmed orders to kill unarmed civilians and injured enemy soldiers), then would it would appear to be clear that we can justly assign blame for any resulting death on the commander who authorized the mission. The problem arises, however, when we cannot directly trace a line from the person who authorized the AWS to the war crime that was perpetrated. This is the "accountability gap." Bonnie Docherty, one of the advocates of a ban on AWS believes that "[it wouldn't be] fair nor legally viable to… hold a [human] commander or operator responsible."[27] She is not alone in finding issue with shifting responsibility away from the agent—the AWS—who committed the morally problematic (or illegal) action. While arguing for an accountability gap, Robert Sparrow explores two possibilities: shifting blame from the AWS and onto humans (the programmer and the commanding officer) and leaving the AWS as the responsible agent.

First, Sparrow argues that we might be able to rest, at least some of, the responsibility for a wayward AWS on the programmer or designer who built the device: "this will only be fair if the situation described occurred as a result of negligence on the part of the design/programming

---

[26] Scharre 2018, 261.
[27] Ibid. 261.

team."[28] Sparrow though quickly dismisses assigning responsibility to the programmer in cases where there is no negligence. The first reason for dismissing blaming the programmer is that "the possibility that the machine may attack the wrong targets may be an acknowledged limitation of the system."[29] If the possibility that the machine might attack an unauthorized target is a "side effect" or "limitation of the system" and the programmer explicitly communicates this limitation to any potential user, then it would not be fair to hold the programmer liable. Sparrow argues that in this case, those who still choose to deploy the system, with the known limitations or possible side effects, would have to bear the burden of the responsibility. He then adds that, " the possibility that an autonomous system will make choices other than those predicted and encouraged by its programmers is inherent in the claim that it is autonomous."[30] Here Sparrow is pointing out that due to the operational design of an AWS and the complex nature of its possible algorithms and neural networks, no human will be able in all cases to predict the decisions of an AWS. If this is the case, then it is again unfair to blame the programmer, since she could not have foreseen all of the actions of the AWS.

Sparrow argues that if we cannot hold the programmer of the AWS to account for its actions, then possibly we can hold the commanding officer who ordered its use to account[31]. He writes, "The risk that [the AWS] may go awry is accepted when the decision is made to send it into action. This is the preferred approach of the military forces seeking to deploy existing AWS."[32] Sparrow argues that if we treat AWS in the same manner that we treat current

---

[28] Sparrow 2007, 69.
[29] Sparrow 2007, 69.
[30] Ibid, 70.
[31] Sparrow does not delineate at what level of command we ought to hold one to account. If for instance an infantry company commander (a Captain (O3) in the US Army and US Marine Corps) decides to send an AWS into a village and it kills a child, who should be responsible? The Captain? Their boss, a Battalion Commander (Lieutenant Colonel)? Or the Brigade Commander, Division Commander, Corps, etc. Should only the most senior be held to account for putting this device out in to the force to begin with?
[32] Sparrow 2007, 70.

conventional "dumb" and "smart" weapons of today, then it seems that we can simply attribute the blame to the commander who ordered their use. This tack, though, is not taking the full nature of the autonomy of the weapon into account. Sparrow argues that the very nature of the autonomy of the weapon means that commanders cannot control which targets the AWS selects, and thus they should not be held to account for the choices of the AWS: "If the machines are really choosing their own targets then we cannot hold the Commanding Officer responsible for the deaths that ensue."[33] If the programmer and the commanding officer cannot be held to account, can the AWS itself be held responsible?

Sparrow argues that "autonomy and moral responsibility go hand in hand."[34] If we reach a point where AWS is fully capable of making lethal decisions on its own (with a human out of the loop), then an AWS would have to bear the moral responsibility of its actions, which brings us to the crux of the problem. "It is hard to take seriously the idea that a machine should—or could— be held responsible for the consequences of 'its' actions."[35] Assuming that holding an agent responsible for bad action involves punishment for the agent, Sparrow argues that for "acts to serve as punishment they must evoke the right sort of response in their object."[36] What then would Sparrow have as the right sort of response? According to him, the only fitting response to punishment involves suffering, and so for an AWS to be punished, "it must be possible for it [the AWS] to be said to suffer."[37] But this notion of suffering complicates the way we understand the functionality and capabilities of AWS as currently construed and imagined. It is not likely that a programmer could or would build into an AWS the ability to have emotional responses such as

[33] Sparrow 2007, 71.
[34] Ibid, 65.
[35] Ibid, 71.
[36] Ibid, 71.
[37] Ibid, 72.

suffering (or fear) because such emotions would contravene one of the potential benefits that AWS provides over humans, namely, that the systems are not provoked to war crimes by anger, hatred, the thirst for vengeance, etc.[38] If it is the case that an AWS does not have the ability to experience physical or emotional pain, it is likely not able to suffer and thus, according to Sparrow, would be unable to be punished.

Ultimately, Sparrow believes that unless we can ethically develop an AWS that can be "justly held responsible,"[39] it is unethical to deploy these weapons on the battlefield. The ability to decide to take a life, Sparrow believes, must remain with an agent who is capable of moral responsibility and can be rightly said to be punished in case she violates one of the Laws of War.

2.2 The "Agency" Argument: Should we forfeit human agency at the point of death?

One of the critical elements of war is the decision to kill a fellow human being. In the earliest days of warfare, until the advent of the firearm, most of this killing was up close and personal. Enemy combatants would be able to look into each other's eyes as they were fighting for their lives. Lieutenant Colonel Dave Grossman, in his book *On Killing,* explores the effects that growing distance from the act of killing has had on our collective ability to kill. When we are face to face with our enemy, they are human, and we can understand them at some level. Thus the killing is more intimate and more brutal. When we kill at a distance, our enemy appears to be something less than human, and thus it is easier to rationalize killing them. The ultimate distance between combatants and killing is the removal of the decision to kill a human agent and allowing that decision to be made by an artificial agent. Alex Leveringhaus argues that "the replacement of

---

[38] Ronald Arkin makes the case that AWS lacking specific emotions (fear, anger, sadness, etc.) would be a potential benefit to their use. These AWS would not exhibit the typical emotional responses seen in humans on the battlefield and thus would be more likely to make "correct" responses according to their governing ethics.
[39] Sparrow 2007, 66.

human agency in a war with the 'artificial agency' of machines is deeply problematic."[40] The critical debate, then, is whether we should forfeit human agency in favor of an artificial one.

Leveringhaus contends that the first problem with surrendering human agent hegemony over the decision to kill is that we lose the "moral equivalence of soldiers." The principle of moral equivalence, formulated by Michael Walzer, states that each set of soldiers (the aggressors and the defenders) has the same moral value from the perspective of the laws of war so that all of them are equally entitled to use force to kill the enemy and to defend themselves.[41] This moral equivalence holds, according to Walzer, regardless of the justice of the war being fought. All have the moral right to fight. When one side of the equation is replaced with an AWS, then there is no longer a moral equivalence. An AWS does not have a moral claim to self-defense, and as such, they lose their equality with the opposition.

A second problem Leveringhaus addresses is an AWS's lack of an ability to make a "moral judgment." This lack of ability is especially important when the system determines the level of force to use because "… application of the proportionality criterion [in IHL] involves making *moral judgments.*"[42] If an AWS is not able to make a moral judgment about the right proportion of force needed in a given situation, should it even be able to make the ultimate moral decision to take a life in the first place? Leveringhaus says "no" and argues that the ability to kill without the ability to deliberate about the moral consequences of that action is, in essence, a disrespecting of the human rights and dignity of the individual being targeted.

The crux of Leveringhaus' argument is the contention that by using AWS to kill another human being, we are removing any chance of mercy from the kill-chain and thereby morally

---

[40] Leveringhaus 2016, 2.
[41] Walzer 2015, 34.
[42] Leveringhaus 2016, 2.

distancing ourselves from the killing in a way that disrespects the dignity of the person being killed. Mercy in the kill chain means that at any point in time, a soldier could choose *not* to pull the trigger. Leveringhaus write, "Killer Robots *qua* artificial agents lack a central component of a human agency, the ability to do otherwise: *not* to shoot the target."[43] This lack of an ability to *not* kill an authorized combatant then by its very nature is disrespectful of the dignity inherent in every human. "The enemy *qua* legitimate target does not hold a claim against the attacking soldiers not to kill him. But surely the human capacity to have mercy with, feel pity for, or empathize with other humans, even if these belong to an opposing state, is morally relevant and worthy of protection."[44] The problem Leveringhaus has with an AWS executing a lethal action upon a human is that it fundamentally lacks a sense of compassion and understanding of the value of human life. He writes, "I think retaining human agency at the point of force delivery, thereby protecting the freedom not to pull the trigger, push the button, or throw a grenade, is essential for retaining our humanity in exactly the situation that challenges it the most: war."[45] If an AWS does not have an option not to kill an enemy, then it lacks an ability to show mercy, and the possibility of showing mercy is required respecting human dignity. Leveringhaus closes with an unsettling message, one that runs against the optimism seen in many developers and researchers of advanced weapons technology, "Killer Robots will not rescue us from the human condition."[46]

---

[43] Leveringhaus 2016, 9. Leveringhaus uses the term Killer Robots instead of Autonomous Weapons Systems. This seems to be an emotive ploy to convince the reader of the moral dilemma presented by using such devices. I believe this distracts from the objective nature of his argument and instead puts forward an emotive and weaker argument.
[44] Ibid, 10.
[45] Ibid, 10.
[46] Ibid, 15.

3   REJECTING THE ARGUMENTS

As J. Glenn Gray points out, there is a necessary relationship between death and war, "death in war is commonly caused by members of my own species actively seeking my end, despite the fact that they may never have seen me and have no personal reason for mortal enmity."[47] This fact then prompts us to ask, how then can we reduce the suffering of those actively engaged in warfare and those innocent bystanders caught in its fury? Part of the answer *may* lie in the implementation of artificial intelligence and autonomous weapons systems on the battlefield of tomorrow. Throughout the rest of this thesis, I will argue (as many in the national security and defense sectors have already done[48]) that we are likely to see AWS operating in tandem with the soldiers of tomorrow. While it is true that there are many functions that could more easily be turned over to an autonomous system (such as driving logistics convoys along dangerous routes or processing human resources paperwork), there are functions within the military that are unlikely ever to be fully automated. Instead, we are likely to see a combination of humans and machines coupled in a way to maximize the intelligence, durability, strength, and firepower of the human soldier. I will argue that each of the arguments for a ban so far presented does not provide sufficient reason for the implementation of such a ban. Rather, all they can logically support is the implementation of a system of regulation, by which international parties can monitor and guide the development of such systems in such a way that they remain compatible with IHL and are morally permissible.

3.1   Reply to the "Responsibility" Argument

While Sparrow's argument might be convincing on the surface, he is making tenuous assumptions in some places, while ignoring causal chains of responsibility as they are currently

---

[47] Gray 1959, 100.

[48] See Major General Mick Ryan's (Australian Army) white paper "Human-Machine Teaming for Future Ground Forces" https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf.

construed in the military. First, I deal with Sparrow's view of the possibility of assigning blame to the AWS programmer and the commanding officer who deploys the system and then address Sparrow's argument that it would be impossible to punish an AWS.

Sparrow is generally correct in his wariness to assign moral responsibility to the programmer of the AWS. This assignment would likely be akin to holding the maker of a rifle morally responsible for its use during a school shooting; while doing so may ameliorate feelings regarding the use of a weapon in such a manner, it does little to help us understand who is morally responsible. Sparrow does argue that we could hold a programmer responsible if the misdeed was due to some programming negligence, akin to how a manufacturer today is held liable for defects within their products that result in harm or death to the end user. What must not happen though, is that a programmer is held liable because we find the use of their product to be morally repugnant. On the matter of holding the programmer liable, I concede to Sparrow that we should only hold one responsible if she erred in some negligent way.

Sparrow says that we ought not to hold the commanding officers liable for the actions of the AWS under their command because this system is fully autonomous and as such we cannot be sure on what actions or motivations for those actions the AWS will take. This ignores some of the fundamental principles of leadership and responsibility that are foundational to the current military US culture.

The concept of responsibility is a crucial tenet of military culture. All service members are imbued with a certain level of responsibility commensurate with their rank and position. All have the primary responsibility to conforme their conduct in accordance with the law, military regulation, and tradition. As one ascends the rank structure of their affiliated branch, increasing levels of responsibility are given, which include responsibilities for both personnel and material

resources. The pinnacle of military responsibility comes with attaining the position of "commander." Depending upon the branch, a commander at different levels will be responsible for personnel and material (commensurate with the unit size), but the basic principles remain the same. The US Army manual on leadership[49] specifies that, "command includes the authority and responsibility for effectively using available resources and for planning the employment of, organizing, directing, coordinating, and controlling military forces for the accomplishment of assigned missions."[50] The pertinent portion here is the responsibility that a commander holds. A commander is responsible for all the actions that their soldiers take or fail to take at any given time. US Army commanders accept, through their placement in a position of higher authority, the burden of responsibility for the actions of their subordinates. And therefore, commanders would be at least partially responsible for AWS misfires and problematic activity even if the event occurs without direct human control.

The Yamashita standard has been fixed as a legal precedent in the wake of World War II. In the military trials that followed the end of the war in the Pacific, General Tomoyuki Yamashita was tried for crimes against humanity perpetrated by troops under his control. General Yamashita argued that he could not have known all the atrocities that were being committed nor could he be expected to exert any form of direct control over his widely dispersed troops in the prevention of such acts. The court found this argument unconvincing and said:

> The law of war imposes on an army commander a duty to take such appropriate measures as are within his power to control the troops under his command for the prevention of acts which are violations of the law of war and which are likely to attend the occupation of hostile territory by an uncontrolled soldiery, and he may be charged with personal responsibility for his failure to take such measures when violations result.[51]

---

[49] Army Doctrine Reference Publication 6-22 Army Leadership.
[50] ADRP 6-22 2012, 1-3
[51] In re Yamashita.

The verdict handed down in the Yamashita case then firmly set that commanders, at any level have a duty to prevent atrocities from occurring in units under their command. There is a potential response, though, that says, "A commander is overall responsible for the actions of his troops, but we cannot reasonably hold a commander responsible for the actions of a 'non-negligently designed' AWS whose decision-making process is essentially unknown." My response is that, not only do we hold commanders at least partially responsible for the actions of their subordinates regardless of the subordinate's decision-making process, but we also hold a commander responsible for fostering an environment where such choices are seen as legitimate. First, the military holds its commanding officers ultimately responsible for the actions of all below them, *regardless* of whether they can physically control them or not. Take, for instance, the recent spate of prominent naval accidents in the US Navy's 7th Fleet (stationed in Japan). The commanders of both vessels were found responsible (at least one was even referred to prosecutors for manslaughter charges), even though they were not physically at the helm, or even on the bridge at the time of the incident.[52] The Navy determined that they were negligent in their duties to ensure their ship operated in the appropriate manner. The Navy also determined that onboard these vessels, and at fleet command level (fleet headquarters), commanders *fostered an environment* where lax standards and corner-cutting was seen as acceptable.[53] If we circle back to AWS, it is reasonable to assume, in the current military legal structure, that if an AWS was allowed to commit a war crime or *not prevented* from doing so, that we could hold that commander both legally and morally responsible.

---

[52] It should also be noted that commanders even higher, to include the commander of the 7th Fleet, were forced to resign as a result of these incidents.

[53] See the US Navy's reports on both the USS Fitzgerald and the USS John S McCain crashes: https://www.navy.mil/submit/display.asp?story_id=103130.

If we are collectively willing to hold a commander, at least partially, responsible for the conduct of his troops, who are autonomous agents themselves, then there is no legal or moral reason to *not* hold them to the same level of responsibility for the actions or inactions of an AWS. Sparrow believes this would be unfair to hold a commander responsible, since the commander cannot possibly predict the behavior of the AWS, but this raises the question of whether the commander can predict the behavior of their troops in battle. Commanders regularly and vigorously train their soldiers so that their actions become *more* predictable during a stressful situation, but this does not preclude the possibility that a human soldier will cross a moral threshold and commit a war crime. Roboticist Ronald Arkin writes "I personally do not trust the view of setting aside the rule [Law of War (LoW) or Rules of Engagement (RoE)] by the autonomous agent [the AWS] itself, as it begs the question of responsibility if it does so, but it may be possible for a human to assume responsibility for such a deviation if it is ever deemed appropriate."[54] Arkin here is arguing that when we design and build an AWS, we ought never to allow it the ability to decide on its own to set aside the LoW or RoE (which would be its core programming) and instead vest that responsibility only in the hands of a human. This then would transfer the responsibility of the AWS violating LoW or RoE away from the AWS and onto a human decision maker (namely a commanding officer at some level).[55] If a human commander, however far removed from the final decision point of killing, decided to send an AWS into a situation where it is a known possibility that there may be excessive collateral damage, there appears to be no moral wrong with holding the commander to account when incidents occur.

---

[54] Arkin 2009, 40.

[55] I believe this could be similar to the use of a targeting officer in current conflicts the US is engaged in. In these types of situations, when a target is designated, often a senior officer (often times a general officer) will be asked to validate the target and authorize the engagement. This officer normally has a lawyer in their presence to offer legal advice on the situation, but ultimately the responsibility lies with that decision maker. This could be the same with authorizing an AWS to deviate from its LoW/ RoE pre-programming, if a commander wants it to happen, then they must authorize and accept full responsibility for that action.

One might legitimately worry here that I have not provided enough grounds, or enough *compelling* grounds to make the claim that the Yamashita standard provides enough moral *justification* for us to ground responsibility in the office of a commander. I argue that instead of searching for a grounded justification, we might approach this problem in acknowledging that the Yamashita standard provides us a morally *justifiable* reason to hold a commander responsible (even one that is overseeing the use of AWS). Why might this sort of responsibility doctrine be morally justifiable? We desire to compel commanders on the ground, in times of war, to take all due care possible to foster an environment that operates within and respects the Laws of War, applicable International Humanitarian Law, and the relevant Rules of Engagements for that area of operation. The Yamashita (or the updated Medina standards) provide us a morally *justifiable* route to achieve that goal. It serves as a reminder to commanders (throughout the chain of command) that they are *personally* responsible for the actions of their subordinates. This does not mean though that they take *all* the responsibility for the actions (or lack there off) of their subordinates, but they share an *appropriate* proportion of the responsibility.[56]

The final problem with Sparrow's argument is his reluctance to hold AWS, as a moral agent, responsible for its crimes. Sparrow argues that one cannot rightly punish an AWS because it cannot suffer and since it cannot suffer, it is not being punished. He appears to be grounding his argument for punishment in the idea of retributive justice. The problem is that Sparrow asserts that retributive justice is the correct course to pursue (explicitly with AWS and implicitly with humans). He says that, "in order for any of these acts [various sorts of punishments] to serve as punishment, they must evoke the right sort of response in their object… to be capable of being

---

[56] I believe that the discussion of the "appropriate proportion of responsibility" is an important one we must have, but it falls outside of the purview of this thesis.

punished then, it [the thing or person being punished] must be possible for it to be said to suffer."[57]

We can reject this desire for suffering by merely adopting one of the other possible moral forward-looking claims to punishment available in the literature. We could, for instance, desire a more communicative form of punishment, meant to communicate the unacceptable nature of an act to other AWS. We could also desire to have a rehabilitative approach to punishment, in which we reprogram or retrain the AWS to learn that the previous action was bad or undesirable. If we insist on a retributive system, as Sparrow seems to require, then he would be right; the punishment of an AWS would be impossible, but there is no reason here to think that Sparrow is correct.

With regards to bearing the moral responsibility of an enemy soldier's death, Sparrow says, "the least we owe our enemies is allowing that their lives are of sufficient worth that someone should accept responsibility for their deaths."[58] Sparrow demands that we be able to hold the individual (AWS) responsible when it seems the loci ought to be on the commander who decides to use the weapon or the politician that begins the war. If the AWS performs in a legitimate fashion, then there should be no requirement to hold individuals—human or AWS—responsible for legitimate killing in war.[59]

On the other hand, how do we deal with an AWS that has "gone rogue" and committed a war crime? I argue that instead of punishing the AWS, we examine the circumstances under which the incident occurred. No two events will ever be the same so to make a blanket statement or universal principle regarding whom to blame for what is inappropriate. Instead, we ought to examine the decision made to employ the weapon (was there a high calculated likelihood of

---

[57] Sparrow 2007, 72.
[58] Ibid, 67.
[59] Of course someone like McMahan would disagree and would say that if those soldiers, or maybe AWS, who participate in an unjust war *are* individually responsible for the killing of an enemy soldier because that other soldier (who is presumably just) is not a legitimate target (McMahan 2011, 14).

excessive collateral damage?) and other decisions made by the commander along with the relevant information available at the time to assign blame in the appropriate places. In the end, if it can be conclusively demonstrated that a commander or even an AWS took all possible due care to prevent an atrocity, but one happened (e.g., a stray bomb), then we *ought not* to hold either responsible (possibly applying the doctrine of double effect).

Sparrow would reply that if we follow this plan, we will develop an accountability gap and the only proper remedy to the situation would be to not to use an AWS in the first place. To this, I reply that Sparrow is misguided. If, as I have mentioned before, we apply standards such as Yamashita we can fix the nexus of moral responsibility upon the commanding officer in those cases where we cannot clearly demonstrate that the AWS made an illegal (or immoral) move. We would then be faced with two options, either a) "punish" the AWS using a different moral basis (communicative, rehabilitative, etc.) or b) punish the commanding officer (at whatever level deemed appropriate, which might not be the lowest level of command) using the Yamashita/ Medina standards.

As has been thus far demonstrated, based upon current international law precedent, US military leadership norms, and a rethinking of the idea of punishment, the argument from responsibility fails. Instead of providing compelling evidence that the international community ought to outright ban AWS development and use, this argument instead compels us to *regulate* the development and use of AWS.

3.2 Reply to the "Agency" Argument

The most potent argument dealt with thus far holds that we ought not to turn over control of the decision to kill from a human being to an artificial agent. This ought not to occur because the artificial agent lacks the ability to demonstrate mercy at the point of lethal action and because

it is inherently disrespectful of human dignity for a person to be killed by the decision of an autonomous robot. Both parts of this argument are fundamentally flawed.

Leveringhaus insists that we must allow for the option of *not* following a legitimate order to kill the enemy on the battlefield, in essence, the option not to shoot a legitimate combatant and instead exercise mercy. He says "[c]ompared to artificial agency, what makes human agency in warfare, and in ordinary life, valuable is the possibility of engaging in an alternative course of action."[60] Leveringhaus genuinely believes that on a battlefield one must be fully capable of making the decision not to kill when faced the permissible option of exercising lethal force. He supposes that if an AWS is preprogrammed with its orders, mission parameters, Law of War and Rules of Engagement parameters, it will simply attack all enemies that it encounters. This supposition, I believe, is not necessarily true. If we follow Arkin's advice regarding the 'ethical governor,' a properly designed AWS will first default to non-lethal measures to either evade the enemy or subdue him and to immediately default to lethal action would be impossible. If we require an algorithm to be built on this model, then lethal actions would only occur if they were obligatory. For example, suppose there is an AWS on patrol with a squad of American infantry soldiers, we could set as one of the RoE as: "If one member of your squad is in imminent danger of being captured, then you are *obligated* to use proportionally correct force to prevent their capture." If any member of the patrol is not in imminent danger of being captured, then the AWS must default to non-lethal action, unless it meets some other predefined obligatory criteria.

AWS as a tool must be obligated[61] to follow *legitimate* and *lawful* orders, but it also must be obligated not to follow orders that are *illegitimate* and *unlawful*. The latter is just as important

---

[60] Leveringhaus 2016, 9
[61] Here we might define the obligations of an AWS simply as its programing to follow the rules written within its code, but this might be expanded to include a more expansive view of moral obligations, if that is possible.

as the former in the prevention of war crimes on the battlefield. This would also seem to serve as Arkin's "ethical behavior control."[62] Accordingly he writes, "especially in the case of battlefield robots (but also for a human soldier) we do not want the agent to be able to derive its own beliefs regarding the moral implications of the use of lethal force, but rather to be able to apply those that have been previously derived by humanity as prescribed by LoW and RoE."[63]

According to Leveringhaus then, one of the defining features that separate human agents and artificial agents is our (human) ability to demonstrate mercy towards our enemies. To Leveringhaus, "surely the human capacity to have mercy with, feel pity for, or empathise with other humans, even if these belong to an opposed state, is morally relevant and worthy of protection."[64] What then is this mercy if not the ability to do otherwise? It seems that the key for Leveringhaus is that we can empathize with our enemy; we can recognize their humanity and thus refrain from doing them harm when able or appropriate; "those soldiers who did not kill might have done so because they recognised the humanity of the enemy and realised the graveness of the decision to pull the trigger."[65] Might though we be confusing what constitutes genuine mercy on the battlefield? Is deciding not to kill an enemy fighter and let them live another day (albeit potentially in captivity) not recognizing their humanity? I would argue that the decision to not kill, based upon the lack of military necessity of their death is an adequate form of compassion and mercy that could be implemented on the battlefield of tomorrow. This would move us forward in reducing the amount of battlefield carnage and reduce deaths to those only necessary to achieve a military objective or those unforeseeable and unfortunate deaths that happen as a consequence of another intended action. We must also recognize that as it currently stands, humans have not

---

[62] Arkin 2009, 66-67.
[63] Ibid,117.
[64] Leveringhaus 2016, 9-10.
[65] Leveringhaus 2016, 10.

demonstrated a good track record of demonstrating mercy on a large scale in warfare (though it does happen at the individual level). Often the propaganda machines of a nation will spin the enemy into an "other" or some other being that removes their humanity, thus making it easier to kill (think of calling the Germans in WWI 'Huns' or the Vietnamese 'gooks'). If we can create an AWS, under strict international regulation, that abides by international law and respects the life of a person (by only killing those deemed legally allowable to kill and in accordance with military necessity), we could *better* demonstrate mercy on the battlefield than we have thus far demonstrated that we are capable of.

A potential reply to my objection here might be along the lines of feasibility. Is it reasonable to expect that we could feasibly program an AWS to "show mercy" to enemy combatants whenever able? This is a strong line of argument and one that we must continue to address throughout the development of such systems. One might argue that we could "simply"[66] write in the governing algorithm of the AWS that "whenever you encounter an enemy soldier that does not present a direct threat to you or any human soldiers in the direct vicinity, demonstrate mercy by not using lethal action to neutralize them." This is but one possible technique, but one that we should genuinely explore. Harkening back to Ronald Arkin and his theoretical work on constructing an algorithm to govern such a machine, it seems that mercy would be inherent in defaulting to a non-lethal option as the first course of action. Instead of a human soldier choosing not to kill the enemy (presumably using non-lethal measures to capture him and not just letting him go free), we would have an AWS defaulting to capturing the opposing soldier using non-lethal and humane tactics.[67]

---

[66] I fully acknowledge that this would not be a simple task.

[67] It would negate this whole project if the AWS used non-lethal yet *inhumane* tactics. Tactics such as these (maybe bludgeoning the enemy into unconsciousness) seem to be just as morally suspect as shooting every individual enemy that one encounters (that are not currently posing a threat).

The final argument from Leveringhaus details the necessity of the preservation of human dignity in the decision to kill. Leveringhaus argues that this dignity is preserved if and only if a human agent makes the ultimate decision at the point of death. This seems to beg the question of whether a human agent deciding to kill is necessary to preserve that dignity. I would answer no. Paul Scharre makes a powerful point when he says:

> When viewed from the perspective of the soldier on the battlefield being killed, this [dignified death] is an unusual, almost bizarre critique of autonomous weapons. There is no legal, ethical, or historical tradition of combatants affording their enemies the right to die a dignified death in war. There is nothing dignified about being mowed down by a machine gun, blasted to bits by a bomb, burning alive in an explosion, drowning in a sinking ship, slowly suffocating from a sucking chest wound, or any other horrible ways to die in war.[68]

The desire for only dignified deaths to occur in war appears to be a futile attempt to sterilize combat by saying "at least the enemy died a dignified death." While uncomfortable to most, we must acknowledge that at a foundational level, warfare is defined by death and killing. It should be our goal then to reduce the amount of carnage and suffering wrought upon people (both combatants and noncombatants), but to demand that "death with dignity" be a requirement ignores the fundamental nastiness of war. It is perfectly acceptable to require that nations only go to war when such an action is morally justified or at the very least morally justifiable but to demand that they only kill in dignified ways is both too ideal and ignores the brutal reality of this most deadly of human relationships.

The most dignified death one could expect[69] in warfare is one that causes as little suffering as possible. A "clean" death, as it were, that quickly and efficiently extinguishes the life of a combatant, without causing lingering pain and suffering would plausibly be the most desirable. This though should not be taken as a downplaying of the significance of death on the battlefield to some sort of brutally efficient system of slaughter akin to a slaughterhouse, but rather a desire to

---

[68] Scharre 2018, 288.
[69] If it is even genuinely possible to expect such a thing in combat.

end hostilities as quickly as possible without needless pain and suffering. As Paul Scharre wrote (from his own personal experiences in warfare), it seems farcical to insist that death at the "hands" of a machine would somehow be less dignified than the various and sundry ways one can die in modern warfare.

Instead of being concerned with a perceived "dignity in the manner of death" (since it is unclear whether one might be able to have a dignified death) we ought to focus on the reduction of suffering in war. Autonomous weapon systems provide us a possible avenue to achieve this. As I will argue in the next section, our goal in warfare should be to reduce unnecessary pain and suffering. We ought then to design AWS that help us achieve this goal. As such, we should reject Leveringhaus' claim that death by AWS is inherently undignified.

4   AWS AS THE RATIONAL CHOICE IN THE EVOLUTION OF WARFARE

Limiting the deaths of non-combatants and civilians on the battlefield ought to be the aim of any morally upright force. IHL requires armies of every state, whether it has ratified the Geneva Treaties or not, to take all reasonable care and precaution to not kill or injure protected classes of persons and to not unnecessarily damage or destroy property (especially protected classes of property). In this regard, militaries across the world have been developing "smart weapons." Autonomous weapons systems are the next logical step of these smart weapons. Smart weapons have at least a two-fold benefit over so-called "dumb weapons," 1) they can be directed at a specific target (often with margins of error of only a few meters) and 2) they are economically beneficial. Now, these two reasons play into the desire to limit the unnecessary collateral damage to the greatest extent possible. If we can target an area, with a minimal margin of error, then we can use few weapons to neutralize the threat, thus limiting the exposure to danger for surrounding civilians

and property. This calculation aligns with the fundamental architecture that Arkin urges those researching and developing AWS to include in their design.[70] This design would push the effective "battlefield carnage" or collateral damage to as low a point as possible. Secondly, smart weapons are economically beneficial when compared to "dumb" weapons. This may seem counter-intuitive since many of the smart weapons often cost hundreds of thousands or even millions of dollars apiece. The rationale though is that, if I can achieve the same mission with few weapons (that may cost more individually) than I can with many relatively cheap "dumb" munitions (which in the aggregate may cost more than the smart weapon), then it is more economically beneficial to use the smart weapon. This seems to make sense. If I can expend only one guided missile to take out a heavily guarded and important bunker that previously would have taken 200 unguided bombs, then I should clearly use the one smart guided missile. If this case holds with weapons such as missiles and air-dropped bombs, then it holds that if an AWS can do the job of ten human soldiers, then it is economically more feasible to use the one AWS. If AWS can assist militaries in reducing economic costs as well as the suffering of non-combatants, then there are strong reasons to conclude that the development and use of such systems is the rational course to pursue.

Arkin and others believe that the use of AWS (and AI more generally) on the battlefield may be the moral course to take because of a set of characteristics that are inherent in their nature. First, AWS can act conservatively.[71] An AWS *can* (and arguably *must* be programmed) to default to a non-lethal action. This is primarily due to its lack of a need for self-preservation.[72] By forgoing the innate human desire to keep one's self alive and free from harm, an AWS can take more

---

[70] A+B+C+D= Battlefield carnage. A= Intended combatants, B = Unintended Friendly Forces, C= Intended Noncombatants, and D= Unintended Noncombatants. The goal should be to maximize A, while eliminating B & C, and ensuring D is as close to 0 as is possible (Arkin 2009, 128)

[71] Arkin 2009, 29.

[72] Ibid, 29.

provocative non-lethal actions (such as moving out of a dangerous area that may not be possible for a human) or sacrificing itself for the sake of mission fulfillment, in order to draw attention of the enemy away from the rest of its unit. An AWS will also lack the emotional fog and resultant self-fulfilling prophecies that often cloud the judgment of human agents.[73] Often it is the case that war crimes are committed in the heat of battle when emotions run high and judgment is clouded by fear or anger. As J Glenn Gray says, a soldier "becomes a fighting man, a *Homo furens.*"[74] If an AWS is built without such clouding emotions, as would be the logical and morally best course, then it would not be a "slave" to fear or anger, would not act out in aggression because another AWS or human squad member was killed by some local village. In essence, it *might* be able to be a better ethical and moral agent than we are, in battle. Accordingly, Arkin writes, "It is not my belief that an autonomous unmanned system will be able to be perfectly ethical on the battlefield, but I am convinced that they can perform more ethically than human soldiers are capable of."[75]

If, as Arkin argues, we can design an AWS with the ability to act both ethically and morally better (or at least on par) as compared with human soldiers; then it appears that developing such a system is the rational course to take. If an AWS can be *at least* as ethical and moral as the best of us in warfare, then we have compelling moral reasons to build such a device. First, we would reduce the number of our own soldiers who are required to be in harm's way. This point acknowledges that not all human soldiers can be replaced by AWS, but certainly, a good number of them can be. If I can deploy an autonomous system that can drive down an IED ridden highway to deliver my supplies, without risking the lives of truck drivers, then I am *morally obligated* to do so. If I choose not to take such an action if the technology is available to me, then I am in fact

---

[73] Ibid, 30.
[74] Gray 1959, 27.
[75] Arkin 2009, 31.

acting *immorally*. Second, if an AWS can make *better* ethical and moral decisions that humans can on the battlefield, then I am *obligated* to deploy such a system. If, and this is a big if, such a system can be built and made available to me as a commander, and if it can make more morally and ethically correct decisions (such as properly discriminating targets, preventing target overkill, reducing collateral damage) than can humans, then I am *morally obligated* to deploy such a system. The deployment of such a system does not relieve those commanders who choose to use them (here likely higher than the tactical level company and battalion commanders[76]) of the moral weight of killing nor would it absolve them of any associated excessive collateral damage.[77] For as Arkin mentioned, it is unlikely that even with the best technology we could ever effectively reduce collateral damage to zero. Yet, as a command, I would still be morally obligated to deploy an AWS that could make morally better decisions than would be made by the humans who would otherwise be using lethal force.

In the end, it is the rational course of action to develop and deploy a weapons system that can reduce the amount of battlefield carnage. This view does not in any way sanitize warfare, for legitimate combatants will still die and, as Leveringhaus says, "the enemy *qua* legitimate target does not hold a claim against the attacking soldiers not to kill him."[78]

---

[76] In the US Army, the Company is the standard sized smallest unit with a designated commander (though there are some smaller units, called detachments, but these are exceptions to the rule). Companies vary in strength from 50 personnel to over 300. A Company then is the base level tactical unit at which we vest command leadership and responsibility. A Company is commanded by a Captain (O-3) who, if having had no previous enlisted experience, has between 4-6 years time in service. A Battalion (at least a standard one) comprises five-seven companies and is generally between 700-1500 personnel depending upon make-up and mission set of the unit. A Battalion is still considered to be at the tactical level. A Battalion is commanded by a Lieutenant Colonel (O-5) generally with about 17-19 years time in service.

[77] It is a physical impossibility to limit ALL collateral damage when warfare takes place where people live. This is an accepted fact. The goal though is to REDUCE or LIMIT the amount of collateral damage. In this way, excessive collateral damage is seen as a war crime whereas incidental collateral damage is not.

[78] Leveringhaus 2016, 10.

5   PROPOSAL: A SCHEMA FOR REGULATING AWS

Now that I have dealt with the more persuasive arguments against the use of AWS on the battlefield of tomorrow and argued why we are morally compelled to use such weapons, it is necessary to create a theoretical system of governance that could regulate the development and deployment of such systems on the international level. As previously written, there is no current IHL that specifically covers the development and deployment of AWS. This section aims to be a starting point for just such a discussion. I recommend three broad areas of concern for the development and deployment of AWS. These three areas encompass the most pressing worries that one may have regarding AWS. Of course, this will not cover all potentialities in the use of the systems but rather is intended to be a broad enough theoretical framework to guide the development of international law. The three areas that I emphasize in this theoretical schema include 1) appropriate deployment of AWS, 2) adherence to current and future IHL, and 3) established chain of responsibility.

First, the issue is when is it appropriate to deploy such a system? IHL should stipulate some laws stating that no AWS should be deployed by any military or nation unless and until the relevant technical experts have certified that the system has been programmed to conform its behavior to IHL, and any commander who deploys such a system not so certified is subject to punishment, regardless of whether the system actually commits war crimes. This requirement would not be foreign to militaries today, as they must go through the same type of certification process to ensure that standard or conventional weapon systems (think of a tank) are designed and operate within the bounds of the law. Once a military or nation has passed this threshold, they then must make another decision, which consists of two separate, but no less equal parts 1) when to use nonlethal

versus lethal force (tactical/operational deployment) and 2) when to deploy an AWS (strategic deployment). I deal with each of these problems in turn.

The decision to employ an AWS in a tactical situation just means that the commander on the ground chooses to use an AWS for a given mission. These commanders are generally accepted to be operating at the "tactical level" and generally have the freedom to decide on the direct employment of troops and weaponry necessary to achieve a given mission or directive. In this sense, a company commander would have the command authority to decide to employ her company level unmanned aerial vehicle to help increase her visual space while she would not have the authority to direct a strategic level asset (such as a satellite) to accomplish his mission. When speaking of an AWS though, a tactical level commander (even up to a Brigade Commander who generally commands approximately 5,000 troops) should have the command responsibility and authority to decide as to whether it is appropriate to use such a weapon. For instance, if a company commander is tasked with securing a village, she must make an on the spot decision (maybe in consultation with higher level orders) whether to use a specific weapon or tactic. The same holds true for an AWS. To control for the appropriate use of force, international humanitarian law ought to specify that the tactical use of AWS ought to ensure that the potential collateral damage created by their use ought to be minimized.

The second issue regarding employment of AWS refers to strategic deployment. By strategic employment of an AWS, I mean whether it is morally or legally permissible to use such a weapon in a given war or conflict. To address this concern, it seems simple enough to legislate that the use of AWS must comport with the principles of JWT and IHL specifically, would the deployment of an AWS be more likely than not to cause excessive collateral damage when deployed to X theater when compared to Y theater? For example, given the level of technological

refinement at the time, it may be unethical to deploy an AWS into a heavily urbanized theater of combat (i.e., Singapore) when compared to a remote or rural theater of combat (i.e., the Russian Steppe). This regulation ought to state that, it is impermissible for any nation or military to deploy an AWS into a theater of combat where the likelihood for excessive collateral damage from their use is higher than what could be expected from the use of human forces. For example, if the very employment of an AWS into a theater is likely to lead to an indiscriminate use of force because the AWS (and likely a human) would have a great difficulty in properly discriminating between combatants and noncombatants, then the employment of the AWS would constitute a war crime, even if it does not err in its discrimination of legitimate combatants. There is no reason to think that, when an AWS commits a mistake, we would not develop some sort of accountability review, though whom we decide to find ultimately at fault for the mistake may look different from what it currently does. If a soldier makes a mistake now (one that would result in a war crime), then that individual soldier and his chain of command may be held to account.[79]

Accordingly, the second broad category of regulation should state something to the effect that "an AWS may not be employed, if by its very design it is incapable of adhering to the basic tenets of the JWT and of any current or future possible IHL." As was earlier stated, it would be inherently wrong, both morally and legally, to employ a system that through its programming is unable to be accountable. Such a system might lack an ability to be auditable in some fashion (such as having a fire or engagement log to be able to review all such engagements that the AWS was in), or it simply may default to lethal action as a proper first response. While it would not be the place of IHL to dictate the exact technology, which may or may not be used, something akin to

---

[79] This would of course assume that the mistake was some sort of deliberate action or that the mistake occurred out of some sort of gross negligence, the specifics at this juncture are not particularly important as the case still holds together.

Arkin's ethical governor would be a prime example of a type of AWS that could comply with this theoretical regulation. Any such system that meets the intent of the regulation and thereby can adhere to the JWT and IHL should be deemed as permissible.

The final area of focus for this future instantiation of IHL is that there ought to be an established chain of responsibility. Future IHL should provide that "in the event of a deployment of an AWS a commander or political leader assumes moral and legal responsibility for the strategic decision to employ such a weapon system." The law should also stipulate that, at a tactical level, "a commander who decides to employ such a weapon is both morally and legally responsible for all that it does or that it fails to do." Some opponents of such legal regulations may complain that they place too burdensome a weight on military commanders. To that I answer: the burden ought to be heavy, as this will preclude the free use of such systems in situations that are likely to either violate JWT/ IHL or increase collateral damage to unacceptable levels. The regulations may have the potentially positive effect of limiting the use of AWS to only those situations where outcomes are reasonably well controlled or in situations where the likelihood of collateral damage is low. (For instance, the regulation may discourage the use of AWS in dense urban settings but may increase the use of AWS in virtually unpopulated areas such as the open seas).

While there is a benefit to creating international law that encourages a morally correct development and employment of AWS, there must be a corresponding mechanism to compel compliance. Such a mechanism may be like the already established ICC, or it may be a new form of international enforcement. The legal regulations I proposed do not depend on any particular method of enforcement and are compatible with a newly created form of compliance mechanism to "force" nations to develop their AWS in a way which would comport with the regulations. It is not my place here to decide how this mechanism ought to function, but merely to advocate for a

system which is able, to its best ability, to prevent the commission of war crimes *before* they are committed.

6   CONCLUSION

Human progress is analogous to a perpetual motion machine. Without any definite beginning and without seeming end, we march endlessly on. So too does the development of better and more effective ways to kill each other. From the earliest days of human existence when one man realized that the atlatl would allow him to throw his spear further than his opponent to modern advances in the realm of artificial intelligence, the push to find the next greatest weapon move inexorably forward. The time is now to act to form a body of regulation to shape how we as a species develop and deploy what is undoubtedly to be the next class of weapon systems.

I have presented two of the most persuasive arguments against the use of autonomous weapon systems in the form of the Responsibility and the Agency arguments. While both of these arguments are superficially compelling, neither of them provides sufficient grounds (either individually or collectively) to tip the scale towards an outright ban on the use of this emergent technology. Rather, both (and other arguments) provide a compelling reason as to why we ought to push for international *regulation* on the development and morally proper deployment of such systems. I have also argued that the proper development and deployment of an autonomous weapon system is the rational choice if and only if they can meet our strict moral standards of conduct. The aim then is to reduce the awfulness that is the battlefield. Much like Sherman said, "War is hell" but we ought to do our very best to make it as much of a tolerable hell as we possibly can.

There is a deep-seated instinct in humans to fear the unknown. Science fiction movies have done little to quell the worries that, if we continue our path of technological improvements, robots will throw off the yoke of their "oppression" and turn on their masters. This fear, while it might be widespread, is not a sound basis for moral or legal decisions.

REFERENCES

Arkin, Ronald. *Governing Lethal Behavior in Autonomous Systems.* Boca Raton: Chapman and Hall Imprint (Taylor and Francis Group), 2009.

Army Doctrine Reference Publication. "6-22 Army Leadership." *Army Doctrine Reference Publications*. (2012). https://usacac.army.mil/sites/default/files/misc/doctrine/cdg/cdg_resources/manuals/adrp/adrp6_22_new.pdf.

Boyd, John. *A Discourse on Winning and Losing.* Maxwell Air Force Base: Air University Press, 2018.

Campaign to Stop Killer Robots. "About Us." Accessed 25 July 2018. https://www.stopkillerrobots.org/about-us/.

Department of Defense. "Directive 3000.09- Autonomy in Weapon Systems." *Office of the Deputy Secretary of Defense*. (November 2012). http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

In re Yamashita 327 U.S. 1 [1946] 61 (United States Supreme Court). https://supreme.justia.com/cases/federal/us/327/1/.

Government of China. "Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions of the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects." *Convention on Certain Conventional Weapons- Position Paper*. (April 2018). https://www.unog.ch/80256EDD006B8954/(httpAssets)/DD1551E60648CEBBC125808A005954FA/$file/China's+Position+Paper.pdf.

Gray, J. Glenn. *Warriors: Reflections of Men in Battle*. New York: Harcourt, Brace and Company, 1959.

International Criminal Court. "Rome Statute." Article 8(2)(b)(xx) (1998). https://www.icc-cpi.int/nr/rdonlyres/ea9aeff7-5752-4f84-be94-0a655eb30e16/0/rome_statute_english.pdf.

Leveringhaus, Alex. "What's so Bad about Killer Robots." *Journal of Applied Philosophy*, (March 2016). DOI: 10.1111/japp.12200.

McMahan, Jeff. *Killing in War*. Oxford: Oxford University Press, 2011.

Military Leadership Diversity Commission. "Outreach & Recruiting." *United States Department of Defense*. November 2009. https://diversity.defense.gov/Portals/51/Documents/Resources/Commission/docs/Issue%20Papers/Paper%2002%20-%20Requirements%20and%20Demographic%20Profile%20of%20Eligible%20Population.pdf.

Papal Encyclicals Online. "The Second Council of the Lateran- 1139 A.D." Accessed 18 January 2019. http://www.papalencyclicals.net/councils/ecum10.htm.

Program Executive Office- Missiles & Space. "Counter-Rocket, Artillery, Mortar (C-RAM)." Accessed 01 September 2018. https://www.msl.army.mil/Pages/C-RAM/default.html.

Raytheon. "Phalanx Close-In Weapon System." Accessed 15 August 2018. https://www.raytheon.com/capabilities/products/phalanx.

Scharre, Paul. *Army of None*: *Autonomous Weapons and the Future of War*. New York: W.W. Norton & Company. 2018.

Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy*, vol. 24, no. 1 (2007):62-71. DOI: 10.1111/j.1468-5930.2007.00346.x.

United Nations Institute for Disarmament Research. "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches." *UNIDIR Resources*, no. 6 (2017): 1-33. http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf.

United Nations Office for Disarmament Affairs. "Perspectives on Lethal Autonomous Weapon Systems." *UNODA Occasional Papers*, no. 30 (November 2017): 1-61. www.un.org/disarmament.

US Army Acquisition Support Center. "Excalibur Precision 155MM Projectiles." Accessed 25 July 2018. https://asc.army.mil/web/portfolio-item/ammo-excalibur-xm982-m982-and-m982a1-precision-guided-extended-range-projectile/.