UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

THE CAMPAIGN TO STOP KILLER ROBOTS: LEGAL AND ETHICAL CHALLENGES POSED BY WEAPONISED ARTIFICIAL INTELLIGENCE AND IMPLICATIONS FOR ARMS CONTROL REGIMES

By

GARETH VAN DER KAAY FARR

Student Number: 23323737

A mini-dissertation submitted in partial fulfilment for the degree

MASTER OF ARTS IN SECURITY STUDIES

In the Department of Political Sciences

Faculty of Humanities

University of Pretoria

Supervisor:

Mr Roland Henwood

September 2021

ETHICS STATEMENT

I hereby declare that I, Gareth van der Kaay Farr (23323737), whose name appears on the title page of this dissertation, have obtained, for the research described in this work, the applicable research ethics approval.

I, Gareth van der Kaay Farr, also declare that I have observed the ethical standards required in terms of the University of Pretoria's code of ethics for researchers and the policy guidelines for responsible research and that the dissertation titled *The Campaign to Stop Killer Robots: Legal and Ethical Challenges Posed by Weaponised Artificial Intelligence and Implications for Arms Control Regimes* has not been submitted by me at this or any other university, that it is my own work in concept and design, and that all material contained herein has been acknowledged.

ABSTRACT

Killer robots invoke scenes that belong in the realm of science fiction. However, weaponised artificial intelligence is making them a real possibility but perhaps not as sentient and evil as science fiction portrays. Weapons that possess autonomous features have been around for a number of decades in anti-material roles but their use in combat directed towards humans has now been documented for the first time. The advancement in artificial intelligence and robotics, especially in the civilian domain, has enabled the development of these weapons. However, how these weapons should be treated within the field of international law is still heavily debated. There are valid arguments to both the legality and illegality of these weapons in the current body of international law with no clear way forward. These weapons also pose complex ethical issues within the moral dimension of warfare, with views ranging from the argument that the development of these weapons is a moral imperative to the argument that they go completely against the morality of war and should be banned. Previous attempts at arms control have yielded successful results with a number of problematic weapons and similar results can be attempted in this instance with civil society leading the charge. This research explores all the dimensions related to the development and use of Lethal Autonomous Weapons Systems from a technological, legal, ethical, and regulatory perspective, concluding that the use of these weapons is likely to increase in the coming years, thus necessitating the urgent development of legal, ethical, and regulatory frameworks to adapt to this new reality.

Key Terms: Killer robots, Lethal Autonomous Weapons Systems, artificial intelligence, International Human Rights Law, arms control, morality of war, ethical robots, machine speed, arms race, just war.

LIST OF FIGURES

LIST OF ABBREVIATIONS

9/11	11 September 2011
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AMRAAM	Advance Medium Range Air-to-Air Missile
AP	Additional Protocol
APL	Anti-Personnel Landmines
ATS	Alien Tort Statute
AUTO-GCAS	Automatic Ground Collision Avoidance System
AWS	Autonomous Weapons Systems
BLW	Blinding Laser Weapons
CBRN	Chemical, Biological, Radiological, and Nuclear
CCW	Convention on Certain Conventional Weapons
CIWS	Close-in-Weapons System
DCT	Divine Command Theory
DoD	Department of Defence
GGE	Group of Governmental Experts
GPS	Global Positioning System
HAF	Haftar Aligned Forces
HRW	Human Rights Watch
ICBL	International Campaign to Ban Landmines
ICC	International Criminal Court
ICJ	International Court of Justice
ICRC	International Committee of the Red Cross
IED	Improvised Explosive Devices
IHL	International Humanitarian Law
IHRL	International Human Rights Law
IR	International Relations
ISR	Intelligence, Surveillance, and Reconnaissance
LASER	Light Amplification by the Simulated Emission of Radiation
LAWS	Lethal Autonomous Weapons Systems
LIDAR	Light Detection and Ranging
NGO	Non-Governmental Organisation

OODA	Observe, Orient, Decide, Act
R&D	Research and Development
RWS	Remote Weapons Systems
SEC	Security and Exchange Commission
UCAV	Unmanned Combat Ariel Vehicles
UDHR	Universal Declaration on Human Rights
UNODA	United Nations Office for Disarmament Affairs
UNSC	United Nations Security Council
UK	United Kingdom
US	United States
USAF	United States Air Force
USV	Unmanned Surface Vessels
UUV	Unmanned Underwater Vehicles
WMD	Weapons of Mass Destruction

TABLE OF CONTENTS

1. CHAPTER ONE: INTRODUCTION

1.1 IDENTIFICATION OF THE RESEARCH THEME

The advancement in Artificial Intelligence (AI) technology has enabled the creation and the possible use of autonomous weapons both by state and non-state actors. With no agreed form of international regulation and the possibility of an incompatibility with current international law, these Lethal Autonomous Weapons Systems (LAWS) represent a new category of scalable and cheap Weapons of Mass Destruction (WMD). As Haner & Garcia (2019: 331) point out, advances in autonomy are increasing incrementally and are flying under the radar with a low political profile. Furthermore, global military spending on AI and LAWS is projected to reach between $18 and $16 billion by 2025. These new weapons have the potential to start a new arms race as well as reduce the cost and time to go to war. They create new ethical and legal challenges with regards to target decisioning and proportionality of force as well as blurring the lines in terms of the accountability for weapons systems' actions. Currently, no form of regulation or agreed standards with regards to the use, manufacture, or development of these weapons systems exists. In October 2012 a number of Non-Governmental Organisations (NGOs) formed The Campaign to Stop Killer Robots, with the mission to have fully autonomous weapons pre-emptively banned (Campaign to Stop Killer Robots 2019).

In November 2012 Human Rights Watch (HRW) (2012), which is also a member of the campaign, published an article in which the authors and the Harvard Law School's International Human Rights Clinic argued that LAWS would be inconsistent with the current body of International Humanitarian Law (IHL) and its deployment could potentially increase the number of casualties amongst the civilian population during times of conflict and that a pre-emptive ban is necessary. LAWS can have a number of negative impacts on civilians, notably erroneous targeting, the reduction in the psychological barriers when initiating a conflict, and the possible lack of assigning accountability and taking responsibility for the actions of the weapons systems (Roff & Danks 2018: 15). This was probably the first time that LAWS came to the public's attention, highlighting the lack of consensus and action with regards to the potential arms control regimes that could be applied to these weapons. In July 2015 experts in

the fields of robotics and AI signed an open letter which called for the banning of LAWS, drawing the comparison between the revolutionary potential of these weapons systems with that of gun powder and nuclear weapons (Altmann & Sauer 2017: 59).

There are indications that AI-powered weapons have already been used and claims have been made that the assassination of the Iranian nuclear Scientist Mohsen Fakhrizadeh was executed on 27 November 2020 with an AI-powered, remotely controlled machine gun by Israel (Kleinman 2020). Israel does indeed possess an autonomous machine gun equipped with a built-in optical system which has been used, and continues to be used, to allegedly kill Palestinians trying to cross from Gaza into Israel. However, it would seem that, at most, it might have been used as a distraction during the attack, enabling human assassins to surprise the convoy (Kirkpatrick *et al.* 2020). However, according to a United Nations Security Council (UNSC) (2021: 17) report, Turkish-made lethal autonomous drones were used in 2020 by Libyan governmental forces against what it deems insurgent forces. These drones were operating fully autonomously and were programmed to hunt down and attack targets without connectivity to the operator.

1.2. PROBLEM STATEMENT

1.2.1. The Legal Challenges

The arguments for the use of autonomous weapons are that they would bestow significant tactical and strategic advantages in battle and, from the moral perspective, that they are preferable to using human combatants due to the possibility of limiting human casualties (Etzioni & Etzioni 2017: 72). However, the use of these weapons poses a number of legal challenges as international law requires a number of precautionary principles on how war is conducted. One of these principles is the concept of applying discrimination in the process of distinguishing between combatants and civilians in the targeting cycle during combat. IHL also requires the evaluation of the necessity of using force as well as the proportionality when applying it. While some decisions can be made before an actual engagement, some decisions need to be made in combat – including the selection of military targets. This is

especially important because there are currently no technical solutions that have been able to incorporate the principle of distinguishing between combatants and civilians during conflict situations (Dahlmann & Dickow 2019: 6). Therefore, a thorough review of the principles of international law pertaining to LAWS would be required.

1.2.2. The Ethical Challenges

From an ethical standpoint, LAWS can be problematic within the international law framework with regards to the idea of human dignity which serves as the basis for both humanitarian and international law (Le Moli 2019: 365). A robot would have no understanding of what it means to kill when it sees a human is a data point in its programme; it does not possess an understanding of what it means to take a life. It also lacks the capacity for reflection on its actions and would therefore violate the victim's dignity. There is also no technical solution currently available that would be able to solve this challenge (Dahlmann & Dickow 2019: 6). However, despite the knowledge of these legal and ethical challenges highlighted above, states are continuing to develop these weapons systems. At the same time, counter arguments have been made that LAWS are essentially a moral good because they reduce human suffering and increase accountability (Müller 2016: 67). As societies have advanced, the ethics underpinning warfare have evolved and, in this context, one needs to ask what impact LAWS would have on the current and future ethical nature of warfare?

1.2.3. Regulatory Regimes

Another question to answer is: should these weapons be regulated? The proponents of regulation also differ in their viewpoints. There are those wanting upstream regulation which is an imposition of the limits on the technological development of such weapons. This is aimed at preventing the conception of the technology needed to make this sort of weapons system possible. Then there are those wanting downstream regulation which responds when necessary to new developments as they occur (Etzioni & Etzioni 2017: 76). Upstream regulation is unlikely to be possible as the technology used in these weapons systems was first developed for civilian applications such as image recognition and other autonomous capabilities as used in

autonomous cars. A further problem is that, even if a regulatory regime is agreed upon, the challenge of enforcement remains.

Regulatory regimes are briefly defined as a complete set of norms, rules, institutions, and actors that are required in order to implement both the process and the outcome of the desired regulation in the targeted sector (Eberlein & Grande 2005: 91). If one examines the high level of compliance with current regulatory regimes it can be argued that this is proof that implementing a regulatory regime does not undermine international cooperation but strengthens it. However, it can also be argued that the reason for this high level of compliance is because governments pre-empt the enforcement challenges and strike shallow agreements. These agreements do not then substantively change the behaviour of governments after entering into the regime (Carrubba 2005: 669). The challenge also remains that past regimes on Anti-Personnel Landmines (APL) and Blinding Laser Weapons (BLW) relied on international NGOs like the International Committee of the Red Cross (ICRC) to monitor and stigmatise violations (Wallach 2017: 29). Therefore, even if a regulatory regime could be conceived, the effectiveness of constraining these types of weapons could be very weak.

1.2.4. Research Question

This study will set out to answer the following research question: with the advent and imminent commercialisation of LAWS, what is the current state of development of these weapons systems, the implications for the security of combatants, non-combatants and civilians and the current challenges for arms control?

1.2.5. Research Aim

This study aims to provide an overview of the current state of development of LAWS, their strategic impact, and what impact this will have on global security and arms control.

Research objective 1*: To present an overview of the current state of the development of LAWS and their impact on security.*

Research objective 2*: To summarise the current positions on the proposed regulation of LAWS.*
Research objective 3*: To assess the impact on the existing arms control provisions.*

1.2.6. Limitations

This study is limited by the fact that this is a developing area within weapons technology. It also focuses on a very broad overview of the many dimensions related to the development and deployment of LAWS.

1.3. LITERATURE REVIEW

In 2008 Noel Sharkey (2008: 1800-1801) wrote an article titled "The ethical frontiers of robotics." As a computer scientist, he argues that service robots have been used successfully in performing dirty, dangerous, and dull work. However, they pose a number of new ethical challenges when performing their duties. This is highlighted in the development of autonomous weapons and the ethical challenge posed by them not being able to distinguish between combatants and civilians in close quarter combat. In order for weapons to be able to do this, computers would need a clear definition of a non-combatant – but this definition would be impossible. They would also not be able to determine when lethal force is inappropriate.

However, more recently LAWS are becoming a very real possibility with advancements in big data and machine learning. As Paul Scharre & Strategic Studies Quarterly (2017: 15-16) highlight, the development of AI has been through a number of hype cycles in the last decades: the initial excitement about the technology leads to hype which then results in disappointment that AI is not able to meet the initial expectations. However, there have been real advancements recently that have made AI a tangible reality and we find ourselves in a period of anticipation and development as concrete progress has been made in the last few years in three dimensions that enable AI. The first is the emergence of big data which is necessary to train the machines. The second is the advancement in computer processing capability that enables parallel computing needed for deep neural networks. The final dimension is

the improvement of the algorithms themselves. This has enabled machines to be effective at solving a number of problems. AI can be viewed as an enabling technology such as electricity rather than a discrete technology. And like electricity, AI will inject life into objects to make them more useful and intelligent. However, at present the application of AI is very narrow and specific to a domain which means that AI does not possess the same general reasoning ability that humans have. Therefore, it remains a very powerful and important technology that is poised to change components of national power in a similar way that past industrial revolutions have transferred power. While AI will shift elements of national power, these new power shifts may look different from the past. Thus, harnessing the power of AI will increase national power and, as a result, some countries have been writing the use of AI into their armed forces strategies.

The then United States (US) Secretary of Defence Chuck Hagel introduced the Defence Innovation Initiative in 2015. This initiative, which became known as the Third Offset, was aimed at maintaining the qualitative military edge of the US over its rivals or near rivals by adopting state of the art technologies into the policy, structure, and operations of the military. Part of the Third Offset's objective is to incorporate the advances in AI to increase the role of autonomy in battle networks and military robotics, ushering in a new era of human-machine collaboration (Leys 2018: 48). The US Department of Defence's (DoD) Unmanned Systems Integrated Roadmap 2013-2038 proposes to not only develop but to use weapons systems with greater levels of automation. The defining feature of this autonomy is the weapons system's ability to identify, select, and engage enemy targets without the need for additional human involvement. While most modern weapons systems have a large degree of computer assistance and more tasks are being automated, the defining feature of autonomy is that the weapons system can think and make decisions about the target selection and engagement (Bode & Huelss 2018: 394).

A wide misconception exists that autonomous weapons are inevitable. However, this ignores that fact that they currently exist and, in some cases, are already deployed (Crootof 2014: 1863). The Israeli Harpy, an anti-radiation drone, is one example of a fully autonomous weapon. No human approves the specific target before the weapon

engages (Scharre 2018a: 62). Engaging a target without human approval is where the legal and ethical issues arise.

Research into the use of LAWS as well as the challenges they pose have been on the increase and a few key trends have been identified in the current research. The first is centred on the ethical and legal implications. Here arguments centre on difficulties in discrimination and the targeting of persons (Sparrow 2016: 97), violation of human dignity (Sharkey 2019: 77), and not meeting the requirements of IHL (Etzioni 2018: 257). The application of IHL strives to realise a fair balance between the military necessity of armed conflict and the humanitarian requirements to protect the innocent. IHL and International Human Rights Law (IHRL) assume that a war's decision makers are human. However, it is becoming clear that, in the future, these decisions will be made either indirectly or directly by machines or networks of machines with varying degrees of human involvement (Saxon 2014: 100), questioning the compliance to IHL and IHRL. If a robot kills a person, there will be no involvement of human thinking with regards to the application of IHL on the proportionality applied or the need to prevent unnecessary suffering, violating Articles 35, 51, and 57 of the Additional Protocol (AP) I of the Geneva Convention (Ulgen 2016: 10).

There is also the challenge of LAWS not having the technical capability of judging if soldiers are *hors de combat*[1], and therefore don't fulfil the requirements of *jus in bello*[2], especially when taking into account the need to preserve the human dignity of enemy soldiers (Sparrow 2016: 112). In addition, the human rights dimension in the case of domestically deployed LAWS also needs to be considered (Heyns 2016: 377). Currently, under these interpretations, fully autonomous weapons could be conceived as being illegal within the current international framework.

The second challenge arises with the need to assign accountability for civilian deaths within the current international law framework. This is an essential component of IHL

[1] Under customary international law, *hors de combat* arises when a person is no longer taking part in hostilities, either by circumstance or by choice. This is achieved if they are incapacitated through injury, sickness, or unconsciousness; if they have been shipwrecked; or if they have a clear intention to surrender (ICRC 2005b).
[2] *Jus in bello* is the body of law that regulates how parties conduct themselves when engaged in armed conflict. IHL and *jus in bello* are synonymous and attempt to minimise suffering (ICRC 2015).

or *jus in bello*. If a weapons system does not allow for the identification of those responsible for the casualties it inflicts it would fail to meet the requirements laid out in *jus in bello* and, as a result, it would not be permissible in war (Etzioni & Etzioni 2017: 75). As such, within the current international law framework, a human remains the legal subject and should ultimately be responsible for the decisions which necessitate the need to be involved in the targeting cycle (Dahlmann & Dickow 2019: 6). This is a critical point in respect of LAWS being able to make their own decisions. If a robot makes a flawed decision it might be very difficult to determine if it was caused by a flaw in the programme or whether it was a deliberate decision made by the weapons system. In this sense, who takes accountability? A clear chain of accountability exists in traditional combat originating from the person who gave the order all the way to the person who pulled the trigger (Etzioni & Etzioni 2017: 75). Therefore, without direct human involvement, LAWS would create an accountability gap and present a challenge to the legal processes to deal with accountability.

A number of arguments have been made for the use and development of LAWS. Arguments have been made that they are essentially a moral good with the potential to reduce human suffering while simultaneously increasing accountability (Müller 2016: 67; Wallach 2017: 29) without violating any current rules of warfare (Leveringhaus 2016: 355). This line of argumentation is explored further with regards to the idea that automation on the battlefield would ultimately reduce the financial burden and human suffering associated with waging war (Steele & Heinze 2014: 98). Robillard (2018: 705) asserts that LAWS are not morally problematic in principle, as it is not possible to separate the decisions and intentions of the weapons system from that of the human decisions and intentions that resulted in development of the weapons systems software.

A number of arguments suggest subjecting these weapons to an arms control regime. Wallach (2017: 28) highlights that large military powers have expressed the potential for a ban and the US has stated publicly that it will pre-emptively support a ban on Autonomous Weapons Systems (AWS) if it is what the member states of the CCW want. However, behind closed doors the large powers have expressed a reluctance to support a ban with many smaller states following their lead. The debate for a ban has been characterised by back-and-forth positions by opponents and proponents. This

should not be interpreted as a stalemate or attempts to convince states that the benefits of LAWS outweigh their potential harms or vice a versa. Those who oppose the ban do not want LAWS to become a special case within in IHL and for those who support the ban the very fact that LAWS pose a unique set of challenges to IHL is their core strategy.

Furthermore, regulating LAWS poses a number of unique challenges with regards to why it should be regulated, what needs to be regulated, and how it should be regulated. Banning LAWS does not easily fit into the traditional models for arms control. In contrast to many of the weapons that have been declared illegal, there are applications of LAWS that would be morally acceptable and, in some cases, even considered a moral necessity, such as substituting LAWS for soldiers and potentially sparing their lives. Unfortunately, this would lower the barriers to entry for armed conflict and some nations would initiate conflict more easily if their political objectives could be met without risking their soldiers' lives (Wallach 2017: 28). Some cite the need to consider the potential positive impact that LAWS could have in fighting unconventional, asymmetric, and irregular warfare that international non-state actors engage in and the impact this has on civilians when considering the banning or regulating of these weapons (Chansoria 2016: 120).

A number of organisations have devoted time to the question of regulating LAWS but no agreement has been made yet. Two distinct bodies apart from The Campaign to Stop Killer Robots have assembled representatives of governments, academia, and industry to deliberate on the increasing automation of warfare, namely the ICRC and the United Nations Office for Disarmament Affairs (UNODA) which is the responsible office for the Convention on Certain Conventional Weapons (CCW). The ICRC has conducted two conferences on AWS while the CCW has held meetings every year since 2014 which have focused on LAWS (Mull 2018: 484-488).

Engaging with research in the legal field will aid greatly in understanding the applicability of the current legal framework as well as its potential shortcomings. Lewis (2015: 1325) wrote an article published in the Yale Law Journal titled "The case for regulating fully autonomous weapons" in which he discusses the legal aspects of LAWS and how a ban is unlikely to be effective, making a pre-emptive ban

unnecessary. He also indicates that the existing principles of international law are satisfactory to define when these weapons can and cannot be used.

In summary through the advances in the underlying technology autonomous weapons have become a real possibility. The US armed forces are planning to increase the levels of autonomy deployed in the forces. However, there are legal and ethical challenges to the deployment of LAWS. These challenges centre on the inability of LAWS to apply IHL and ethical considerations in their current form due to the technical limitations with the current state of the technology. There have been some arguments made that these weapons would be moral good however, are a number of organisations dedicated to bringing these issues to the attention of the public to influence the creation of an arms control regime aimed at banning the development and use of these weapons.

1.4. RESEARCH METHODOLOGY

This study makes use of a qualitative research approach focused on examining current literature. Creswell (2014: 4) highlights that it is an approach for investigating and understanding the meaning that individuals or groups (in this case scholars and practitioners of the contributing disciplines) assign to a human or social problem (in this case the development of LAWS). It is an approach that comprises of developing questions and evolving techniques with the data being compiled in the form of words. Knowledge is produced inductively by constructing general themes from underlying particulars, with the researcher then interpreting the data. This approach is pertinent to my research as I will be examining the literature across a number of disciplines in an attempt to draw out the main themes in order to determine whether or not consensus exists with regards to the current landscape of research into LAWS and to summarise potential approaches to arms control.

The proposed research will be grounded in normative theory. Normative theory in International Relations (IR) is concerned with the moral dimensions of the discipline and is described as the moral philosophy of IR. Normative theorists see their theory as being both about facts and values. The facts can be defined as the areas of the

discipline that contain normative content, such as the practices, institutions, or rules that have been established. A pertinent example is the rules about the conduct of war. Normative theorists wish to offer a theoretical explanation of said normative practices, institutions, and rules (Jackson & Sorensen 2003: 260). As this research will be examining the rise of LAWS it presents an opportunity to explore the potential impact on the rules of how war is conducted. It would thus fall squarely in the normative domain. At their very core, the arguments for and against LAWS are moral and centre on the taking and preserving of human life in times of war. As normative theory in IR is concerned with moral aspects of the discipline as well as the rules used in the international area, it is best placed as the theoretical foundation of the proposed research.

The most appropriate approach to examining this inquiry is through a scoping review. A scoping review attempts to rapidly map key concepts currently underpinning the available research across the disciplines (Arksey & O'Malley 2005: 21). A scoping literature review is useful in providing an initial assessment of the potential size and scope of the research literature from a variety of disciplines currently available, including on-going research (Grant & Booth 2009: 101). This design will be useful to explicitly highlight where convergence, divergence, or consensus exists across the disciplines around the idea for the need and implementation of regulation for LAWS.

An additional benefit of scoping reviews is that they set the scene for future research by critically reviewing the available research to identify the knowledge gaps that will help point the way to refining questions, concepts, and theories to be studied in future research (Jesson *et al.* 2011: 76). This is applicable to the proposed research in that the research into LAWS is evolving and there is not much in the form of a synthesis of ideas across disciplines. A scoping review is best placed to assess and arrive at potential knowledge gaps within the interdisciplinary literature on the appropriateness and challenges of developing a regulatory regime for LAWS.

The proposed research will mainly consult secondary sources. The use of secondary sources supports the objectives of the scoping review because they bring to light the reuse of already produced insights to develop new social scientific and/or methodological understandings (Irwin 2013: 295). As secondary sources represent

the most coherent and up to date knowledge produced on the topic within the respective disciplines, it represents the best base to begin to tackle the objectives of this proposed research – which is to articulate the themes (whether common or divergent) across the disciplines and to examine the feasibility of the implementation of a regulatory regime. I will be examining a range of secondary sources, including journals of the respective disciplines; conference papers such as those produced for the International Research Conference Robophilosophy 2018 held in Vienna, Austria; news agency sources such as the Economist; as well as websites of organisations, for example, the Campaign to Stop Killer Robots and the ICRC. I will be examining texts across the disciplines to find arguments which have a position either for or against regulation in some form to discover whether or not they favour regulation and what form this takes.

The proposed research is going to use a thematic analysis to identify recurring topics, ideas, or themes in the literature. Thematic analysis is a method used to recognise patterns in that the themes that emerge from the literature become the categories for analysis (Roberts *et al*. 2019: 1). It can be used to uncover issues, problems, similarities, and differences within the themes (Hawkins 2017: 1757). This is achieved through a careful reading and rereading of the data (Fereday & Muir-Cochrane 2006: 81). A thematic analysis can be useful when investigating a phenomenon where little prior understanding exists or where an understanding is generated from multiple and diverse perspectives. As indicated, LAWS is not a new phenomenon, and the legal and ethical challenges they pose are unique. The need to develop (or not to develop) a regime to regulate them is a normative imperative. Therefore, this type of analysis suits the planned research as it will enable the rapid identification of similarities and differences across the various disciplines of computer science, international law, and IR. This will enable the development of a comprehensive understanding of the various viewpoints that exist across the literature of this emerging problem.

1.5. RESEARCH STRUCTURE

In exploring the current landscape of autonomous weapons and the implications for arms control, the research will be structured into five chapters and will be presented

as follows: chapter one will deal with the introduction, framing of the problem, and the research overview. Chapter two will explore the first research question examining the current state of LAWS with regards to their development and the subsequent military applications. Chapter three will address the second research question through examining the current trends in the research and the summary of the current attempts at a regulatory framework. Chapter four will address the final research question on the current arms control landscape and insights that could be useful for LAWS. Chapter five will summarise the ethical issues with LAWS and chapter six will conclude the research by summarising the study and highlighting the findings.

1.6. ETHICAL CONSIDERATIONS

There are no ethical concerns that will be encountered in the completion of this study. The proposed research will use data that is available in the public domain. The study will adhere to the ethical guidelines as determined by the faculty, and ethics clearance will be sought from the faculty research ethics committee. The proposed research will also adhere to the University of Pretoria's plagiarism guidelines.

2. CHAPTER TWO: EXPLORING THE BACKGROUND AND CURRENT STATE OF LETHAL AUTONOMOUS WEAPONS

2.1 INTRODUCTION

With the continuous development of AI, autonomy in weapons is perhaps the biggest threat to the stability of the international system since the Cold War. This chapter aims to explore and define what autonomous weapons are and clarify what makes them unique as weapons systems. It will give an overview of the current state of development of LAWS. It will also examine the underlying technology of AI and how this will power these weapons systems. It will outline the different forms of autonomy and how they relate to humans in terms of control. There will also be a discussion about the incentives for militaries in adopting autonomy and will aim to clarify whether these weapons systems already exist and what this means for a potential arms race.

2.2 WHAT ARE AUTONOMOUS WEAPONS?

Policy makers and scholars seldom agree on the exact definition of what makes an autonomous weapon. This is further complicated by the different understandings that scholars and military commanders have about what these weapons systems can and cannot do. This then leaves the door open for people to use the available science fiction analogies of the robot assassin or the rise of Terminator. Not correctly defining AWS could lead to poor strategy and inadequate policy making (Leys 2018: 50). The most widely used definition of AWS originated in the US DoD directive number 3000.09 which defines an autonomous weapons system as:

> *A weapons system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapons systems that are designed to allow human operators to override operation of the weapons system but can select and engage targets without further human input after activation* (United States 2012: 13).

For the ICRC, AWS are weapons systems that can learn from their environment and in response adapt its functioning. An AWS would be able to search for targets, identify them, and engage them with lethal force if required without any humans involved in the process (ICRC 2013). One can think of the automation of these weapons systems or machines in general as giving them the ability to sense their environment, think about a course of action, and execute the desired action. The level of independence of each of these actions will fall on a spectrum of autonomy (Thomas 2015: 240). Therefore, autonomy in LAWS is a continuum and can be applied to distinct functionalities of the weapons system.

How autonomous a weapons system is can be broadly classified into four dimensions of autonomy, progressing from inert weapons to automated weapons, semi-autonomous weapons, and finally to autonomous weapons.

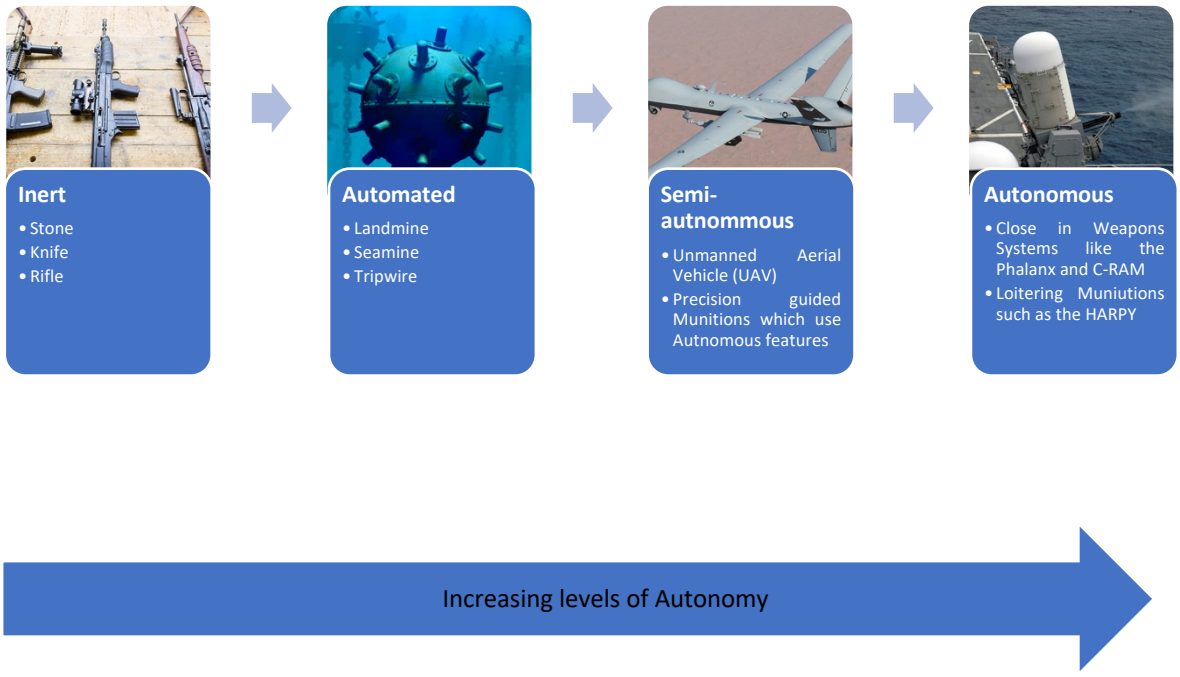

Figure 1: Spectrum of autonomy in weapons systems (Crootof 2014: 1864)

Inert weapons are classified as weapons that need synchronous use by a human to kill. Examples include anything from a stone or a knife to a rifle. Automated weapons are weapons that are exclusively reactive and are often deployed ahead of any

potential engagement with a target. These weapons merely follow what they are programmed to do with no ability to interpret gathered data or run decisioning algorithms to make its own decisions about how to react. For example, a landmine, once triggered, will explode regardless of what triggered it, whether it is an enemy soldier, livestock, or even a child. Automated and autonomous weapons could react in a predictable manner to certain situations but, when triggered, automated weapons do not have a choice in their course of action. Semi-autonomous weapons possess a few autonomous capabilities. Examples include the ability to identify a target, select a target, and possibly the choice of how to engage the target. However, all of these automated capabilities require human approval of the action to be taken. Autonomous weapons, on the other hand, would be free and capable to select and engage targets based on information gathered and within programmed constraints but without the need for any human decision support (Crootof 2014: 1864-1865).

There is also the consideration of the weapon used and whether it is classified as autonomous or not, as in the example of the Advance Medium Range Air-to-Air Missile (AMRAAM). The AMRAAM missile is fired by pilots using long range radar data with no visual confirmation of the target. Once the missile is fired it uses its own internal radar and navigation to locate and destroy the target. In this context, if the pilot decides on a valid target and has chosen to select and fire the AMRAAM, it is a semi-autonomous weapon. If these semi-autonomous weapons are used in accordance with the rules of warfare which have been designed by a responsible military then they cannot be considered autonomous. However, in the case where a pilot is unsure about the target and fires multiple missiles in the vicinity of multiple aircraft in a wider area, the missiles would potentially lock onto any one of them and subsequently destroy it. In this case the pilot would not know what aircraft would be destroyed and, depending on when the seeking mechanism turns on, each of the aircraft in the vicinity may be attacked or one single aircraft attacked multiple times. In this instance, the argument can be made that the missiles are making choices and could thus be classified as autonomous weapons (Horowitz 2016a: 93). For this reason, it is essential to contemplate the context of the application of the weapons system in order to define the potential autonomy thereof.

2.3 WHAT IS ARTIFICIAL INTELLIGENCE AND HOW DOES IT ENABLE LETHAL AUTONOMOUS WEAPONS?

It is impossible to have a discussion about LAWS without a discussion of the underlying technologies that make these weapons possible. The current speed of technological progress is continuously opening up avenues for the incorporation of autonomy into many systems, including weapons. Most of the advances in the discipline of computer science and robotics which are crucial to the advancement of autonomy in weapons are actually dual use technologies that are driven by automotive or communication and information industries. The underpinning technology for autonomy in machines is AI (Haas & Fischer 2017: 287). Currently, AI is potentially one of the most financially lucrative scientific fields and has been said to be the fuel for the fourth industrial revolution (Surber 2018: 2). As Horowitz (2019: 767) explains, AI will be an enabling technology – much like electricity – and the potential military applications of AI will be wide-reaching, including both kinetic weapons like LAWS as well as non-kinetic weapons with example of AI being used in cyber warfare.

When trying to explain the use of AI within the context of weapons it becomes challenging because the term contains the word “intelligence”, a traditionally human trait. In spite of a lengthy history of both research and deliberation, no standardised, accepted definition or a general understanding of this human trait exists. Therefore, it is necessary to define what is meant by ‘intelligence’, especially when these AI systems will be very different in nature to humans and their natural cognitive abilities (Surber 2018: 2). Legg & Hutter (2007: 402) define it as follows: “Intelligence measures an agent’s ability to achieve goals in a wide range of environments”. This would include the capacity to learn, understand, and adapt which would enable the capacity for problem solving in varying environments. The above definition is applicable to both humans and machines. However, what makes it artificial and what impact does this have?

It is essential to explain what is meant by AI and what the distinctions are. AI can be conceptualised as two discrete but interconnected definitions, and while these might not be two meanings with universal validity, it should provide some insight to the danger of anthropomorphising machines, especially in the realm of autonomy and

intelligence. While it may seem logical to describe these machines using this traditionally human terminology, especially when the machines are increasingly able to execute tasks that resemble human behaviour, this would further complicate the true understating of what it means to have an intelligent and autonomous machine. AI can be understood as a scientific discipline with intellectual roots in Greek mythology whose contemporary history began with stored-program computers. As a scientific discipline, AI is the attempt to find out how the human brain enables feelings and thoughts. AI started with the idea that learning and other features of intelligence can be accurately described, enabling the possibly of a machine being able to simulate it. Consequently, the study of AI refers to the investigation of the computations that would enable a machine to perceive, reason, and act. It also refers to the work required to make machines think and the enabling of machines to perform tasks that would otherwise require an intelligent human (Surber 2018: 3).

Surber (2018: 4) defines it as “a scientific undertaking that is aiming to create software or machines that exhibit traits that resemble human reasoning, problem-solving, perception, learning, planning, and/ or knowledge”.

The main components of AI research comprise of the following: knowledge engineering, which focuses on providing machines or software with rich knowledge of the world; machine learning, which studies the use of algorithms applied to insights gathered from data to predict outcome; reinforcement learning, a sub-discipline of machine learning, which focuses on algorithms that learn from trial and error; deep learning, which is machine learning that uses neural networks and is behind breakthroughs in speech recognition and computer vision; machine perception, which uses inputs from sensors to gather aspects from the environment it finds itself in; computer vision, which focuses on the analysis of visual inputs; and, finally, robotics with a focus on both robots and the computer programmes that control them (Surber 2018: 4).

AI can also be referred to as the ability of software or hardware to perform tasks as a result of the research on AI as detailed above (Surber 2018: 4). This enables the imitation of human thought and perception, giving machines the capability to execute complex tasks, including decision-making, which before was limited to humans (Hass

& Fischer 2017: 286-287). Put another way, AI simulates the human cognitive function solving a set of specific defined problems and tasks (Ruta 2018). These abilities, like driving a car without human input or recognising speech, can be understood as artificially created intelligence (Surber 2018 :4). Therefore, AI can be understood as the harnessing of computing power to execute an algorithm in order to perform a task that previously needed human intelligence (Horowitz 2019: 767), while simultaneously creating a commodity which results in AI being both a possible source of wealth as well as the means for political influence (Surber 2018: 4).

As such, AI could secondly be conceptually defined as "the formless capacity embedded in software and hardware architecture which enables the latter to exhibit traits that resemble human reasoning, problem-solving, perception, learning, planning, and/ or knowledge" (Surber 2018: 4).

There is, however, a further distinction that is drawn between different types of AI. AI applications are divided into two subsets: the first is narrow or weak AI and the second is strong AI, or what is also known as Artificial General Intelligence (AGI). The second definition refers to narrow or weak AI (Surber 2018: 5) which is generally slightly faster at performing tasks than a human would be. In other words, the range of intelligence and applicability is discrete. An example of narrow AI in practice is finding patterns in large volumes of data, recognising text on an application form and automatically populating the relevant data fields, or only driving a car. AI trained at image classification would not be able to perform language translation without being retrained. On the other hand, AGI would be better than humans at general cognitive tasks such as strategising, economic productivity, and social manipulation (Ruta 2018; Haas & Fischer 2017: 287).

The creation of AGI is the long-term goal of many researchers and can be defined as the intelligence capacity embedded into systems which approximate the human mind. If this goal is achieved, it might be possible to conceive artificial superintelligence which is the creation of an intellect that will far exceed the cognitive function of humans in all domains of significance (Surber 2018: 5). Therefore, AI does not have a universal definition and is currently far away from the science fiction depiction of what current AI can do.

When applying the above definitions to the creation AWS, narrow AI technologies would be used. The way that intelligence would be generated in these weapons systems would be through the use of almost all of the AI disciplines detailed above with a focus on machine learning and its sub-disciplines as well as computer perception, computer vision, and robotics. Therefore, the AI in an AWS would be the result of training the underlying algorithms repeatedly using large volumes of data. In the context of image recognition on the battlefield for an autonomous tank, the algorithm that is responsible for the detection and acquisition of a target would have been fed a large volume of images of different models of tanks in various positions, lighting, and weather conditions in order for the algorithm to learn reliably what a tank is, what an enemy tank is, and what neither is. This is much the same way that human babies are taught and learn to differentiate between different objects. The AI would also have to be taught what the significance of identifying a tank is and what its next course of action should be. The autonomy of the weapon is generated from the ability to recognise its environment, through the training of the underlying algorithms, and then to make a decision with regards to what action to take within a number of predetermined actions based on the information it understands. Finally, it is enabled to take the necessary course of action.

The traditional approach to computer-run systems has been that they are programmed by humans following a rules-based approach. In other words, the programmer delineates the problem to be solved as well as how to solve it. This approach is suited for tasks that can be clearly defined and limited to a small number of variables that are unknown but can be mathematically described. This ultimately results in software that will be limited by the capability of the programmers to model highly complex environments such as a battlefield. Battlefields are complex and nonlinear. However, as described above, new techniques such as machine learning are emerging which allow the machine to learn and adapt from its own experiences rather than to programme them to think in a predefined manner. The additional subtype of machine learning is deep learning, which uses artificial neural networks similar to those found in the human brain. This technique is described as one of the enabling technological advancements that would provide an advantage on battlefields of the future. Machines that use deep learning would potentially be able to operate in highly complex

environments and would be able to adapt very quickly to that changing environment. This ability would be essential in an environment where milliseconds count and errors could have grave consequences. Deep learning is powered by the vast and speedy analysis of data far beyond what a human can process. These algorithms are trained on large volumes of data and, once they are deployed, they learn from their experience (Haas & Fischer 2017: 287). Thus, the resulting weapons system utilising the above techniques could prove to be a very powerful weapon and thus highly desirable for both military and non-military actors.

Even though deep learning is a powerful technique, AWS will not have AGI; therefore, the weapons will not possess traits generally associated with humans such as emotion, consciousness, or the broad semantic understating that is necessary for decision-making with a moral dimension (Noone & Noone 2015: 27). This results in a further challenge in understanding why the machine behaved in the way it did if something went wrong. For example, when training a machine, the input with regards to the training data set and the resulting output are traceable. However, the process used in learning from the data is not and therefore this process could be best described as a black box. The black box problem stems from AI's use of machine learning algorithms that use data in ways that are not readily understood or auditable by humans. This black box problem may arise from two dimensions. First, a lack of transparency is generated due to the complexity of the algorithm's structure, for example when utilising a deep neural network that uses thousands of artificial neurons which work collectively in dispersed ways in order to solve a problem. This example is known as the complexity black box. The second is that the lack of transparency could be created as a result of the AI utilising an algorithm in a machine learning approach that uses geometric relationships that human beings are unable to visualise. This example is referred to as the dimensionally black box (Bathaee 2018: 901).

These issues make the assessment of exactly what capability the machine has developed and how it will react in a situation with great certainty extremely difficult (Haas & Fischer 2017: 288). As an example, researchers at the University of Washington undertook the task of distinguishing between wolves and Eskimo dogs (huskies). An image classifier was trained on 20 hand-selected images that all pictured wolves with snow in the background while the pictures of the huskies did not picture

any snow. The model produced the output “Wolf” when the image had snow or a light-coloured background in it and “Husky” for all other images, regardless of the position, pose, animal, colour, etc. Only the ground with the snow cover was given as an explanation for the classification (Ribeiro *et al*. 2016: 8-9). This goes to show, it is not always possible to understand exactly what the machine will use to learn which will result in less than certain knowledge as to how the machine will behave.

AI is the key technology that underpins the possibility of LAWS and while AGI might have us thinking in apocalyptic terms, it is the more muted narrow AI that will power autonomous weapons. Even so, narrow AI can be complex and arriving at an understanding for why the system made a simple decision can be incredibly difficult but the foundation for building LAWS using current AI capabilities has already been laid.

2.4. OBSERVE, ORIENT, DECIDE, ACT

To aid in the explanation of the difference between automation and autonomy, it is crucial to understand how the decision-making process in these systems operate. The Observe, Orient, Decide, Act cycle, or the OODA loop as it has come to be known, is a useful tool for understanding complex systems. When discussion takes place about autonomous weapons and if humans are “in the loop”, “on the loop”, or “out of the loop”, it is the OODA loop which is being referred to. It was invented by Colonel John Boyd who was a United States Air Force (USAF) pilot and military strategist. He felt that when pilots engage in a dogfight, the advantage lay with the pilot that can make decisions faster and more accurately than his opponent, thus disrupting his opponent’s decision-making (Marra & McNeil 2013: 1144-1145). A simple explanation is that one first observes the action the enemy is taking, orientates themselves to the enemy action, makes a decision based on the enemy action, and then takes their own action based on that decision. The military believes that speed is the most important component of the loop and whoever can cycle through the loop the quickest will succeed (Coram 2002: 272-275). Boyd believed that when a commander has developed the appropriate deep intuitive understanding of a changing situation, the tempo of battle will pick up and the commander will be able to bypass the Orientation

and Decision part of the loop and be able to Observe and Act almost simultaneously. This adaptability gives the OODA loop its power. In turn, a deep understanding of the OODA loop gives the commander the ability to compress time between observing the situation and taking action (Coram 2002: 275). Therefore, in the context of the OODA loop, automation would refer to automating the individual components that make up the OODA loop such as observing. For LAWS to be truly autonomous on the battlefield, it would have to execute the OODA loop without human involvement.

2.5. AUTONOMY

A further challenge to the discussion about autonomous weapons lies in the definition of autonomy. The dialogue can get derailed attempting to apply the understanding of human autonomy to a robotic system during philosophical debates (Toscano 2015: 193). As Noorman & Johnson (2014: 52) highlight, human autonomy generally suggests a human's ability to act on their own, control themselves, and assume responsibility for their actions. However, the autonomy of a machine could be classified on a sliding scale which ranges from where humans have direct control to that of adaptive capability where the machine can adapt its functionality to that of the changing environment it finds itself in and take action (Boogaard 2015: 251).

Boogaard (2015: 251) proposes a four-dimensional approach to the measurement of the level of autonomy in a machine. The first is how frequently the system needs to interact with a human. The second is the how well the system tolerates environmental uncertainty. The third is the level of discretion it has when changing its plan and the execution of its task without having to further involve a human. The final approach is the system's ability to learn from the actions that it took and add new potential avenues of action to its ever-evolving inventory.

Scharre (2018a: 40) presents three dimensions for measuring autonomy. The first is the type of task that is being performed by the machine. The second is the connection between the machine and the human operating it during the execution of the mission and the third is the level of sophistication of the decision-making capability of the

machine while executing the task. The autonomy of a machine can be increased if any one of these dimensions is increased.

Consider the first dimension, namely the type of task that the machine is performing, as tasks are not equal in complexity, significance, and risk. For example, a thermostat is an autonomous system regulating temperature. Machines will often perform some tasks autonomously while leaving the human in control of other tasks, leading to a blend in the control of the system. A good example would be that of a modern automobile with numerous autonomous features such as anti-lock brakes or adaptive cruise control, but a human is still in charge of steering. Airbags are an example of an autonomous system that is always ready to go with the sole discretion of deciding when to deploy. There are also systems that are designed to specifically take control away from a human in certain situations. An example of this would be that of US fighter aircraft that has been fitted with an Automatic Ground Collision Avoidance System (Auto-GCAS). In this example, the aircraft systems will override the pilot inputs or take complete control of the aircraft in the instance that the pilot loses consciousness to avoid a collision. These systems demonstrate that it is not possible to discuss autonomous systems without understanding which task is being automated.

In returning to the example of the automobile, humans are still driving them but an increasing number of autonomous systems are assisting the driver, or even controlling the vehicle for a brief period of time. The machine becomes more autonomous as it takes on more tasks but some degree of human involvement remains. A fully autonomous car would be able to navigate and drive by itself, but a human would still be selecting where to go (Scharre 2018a: 41-43). Therefore, it is crucial to understand that you could have a weapon that has autonomous target identification and selection but requires a human to engage the target or vice a versa. There are many dimensions of a weapons system that could be automated but only when all functions are performed autonomously would you have a true AWS.

The second dimension is that of the human-machine relationship which consists of three types of interactions. Autonomous systems perform a cycle similar to that of the OODA loop discussed above. Systems can be classified by where the human fits into the OODA loop. The traditional human “in the loop” cycle requires a human to

purposefully engage a target. The “on the loop” cycle does not need a human to purposefully engage the target; however, a human can intervene and stop the engagement. An “out of the loop” system requires no human input or intervention in the engagement of a target (Hauptman 2013: 170). In this context, a semi-autonomous operation can also be called a human in the loop system which compresses the observe and orient stage into a sensing stage, thus operating in a sense, decide, act loop. In semiautonomous systems a human can interrupt that loop with the machine focused on the sensing of the environment and decide on a course of action, leaving the human to act. The second is the supervised autonomous operation also known as human on the loop. In this case the system performs the sense, decide, act loop without human involvement but a human is able to intervene at any time and stop the action if necessary. The final category is a fully autonomous operation, or human out of the loop. In this scenario, once the machine is activated, it senses, decides, and acts on its own without any feedback to or from the human (Scharre 2018a: 41-43).

The third dimension is that of the intelligence of the machine. The more intelligent the machine, the more complex the task, and the more challenging the environment that the machine can handle. One can distil the spectrum of intelligence found in machines to three categories: automatic, automated, and autonomous. Automatic systems do not have much decision-making capacity. They sense the environment and then act accordingly. An example would be an automatic light switch used for streetlights. When the system senses darkness, it switches on the lights and when it senses light, it switches them off, resulting in an extremely predictable system. On the other hand, automated systems are more intricate in that they asses a number of inputs and ranges before taking some form of action. A modern engine management system which optimises the combustion process is an example of this. Given the inputs and range of programmed parameters, the behaviour of the system could be anticipated by an experienced user. Autonomous systems, on the hand, have complex internal functions, and while the user will understand the task that needs to be performed, there will be little understanding of the how the system performs the task. This is referred to as being goal-orientated, with the human specifying the outcome, but the system having the flexibility of arriving at the outcome (Scharre 2018a: 44-45).

True autonomous weapons would not have direct human control of their actions. These weapons systems will execute their programmes and, unlike humans, exercise no free will or self-determination, as algorithms are now able to automate various functions within a system. It is this automation of processes that make a machine autonomous and, in the case of LAWS, the decision to kill would be made by an automated process which would consist of an algorithm defining targets (Asaro 2019: 539). Therefore, when examining LAWS, one needs to consider the type of task being automated, the human relationship to the automation of the task, and how intelligent the system is. For a system to be truly autonomous, it would need to have no human involvement in any of the enabling tasks it performs.

2.6. MACHINE SPEED

It is important to introduce the concept of machine speed. A distinctive characteristic of LAWS, in addition to their independence, is the speed at which these weapons systems can gather and interpret data and execute decisions based on that analysis far beyond human cognition. This speed will convey substantial military advantage (Saxon 2014: 104) and the pressure to remove the human from this cycle will grow as the operational advantage of the speed at which LAWS can operate becomes apparent in comparison to the slow response time of humans (Bode & Huelss 2018: 398). In fact, combat of the future could be conducted at such speed and with such complexity that human operators might become unable to keep pace and reliably direct operations. In this scenario, if a force does not employ some autonomous capability, they would be operating outside of the enemy's OODA loop and would thus be handing the initiative to the enemy (Schmitt & Thurnher 2013: 238-239). The emphasis here is on the speed of completion of the OODA loop. If this is achieved, victory will go to the side that finishes the OODA loop the quickest (Scharre 2018a: 34). The USAF concluded that computing speed and capacity advancements directly impact the speed at which the OODA loop can be completed. This would directly shift the role that computers have in the battlefield from that of support to participating in full with humans in completion of the OODA loop. Estimates are that by 2047 the OODA loop will be completed in micro or even nanoseconds, and will inevitably shift humans from being in the loop to being on the loop (United States Air Force 2009: 41).

There is little uncertainty that in the future of unmanned weapons systems speed will be the most important factor, regardless of the domain of the weapon. In this future, humans can play no part. Even the US has conceded that lethal authority may inevitably need to be delegated to machines due to the need for speed. This is already envisioned in the domains of electronic and cyber warfare as delegation is necessary when humans no longer possess adequate reaction times. This concession becomes even more likely if an adversary concedes more control to their LAWS which would necessitate the replacement of humans just to be able to keep up in a potential battle (Altmann & Sauer 2017: 124).

Machine speed is a revolutionary concept on the battlefield and is likely to become one of the key issues around LAWS. This aspect alone could drive an arms race as the ability to control a fighting force capable of operating at a pace of battle faster and with more endurance than any human force is capable of would be a great strategic advantage.

2.7. INCENTIVES TO ADOPT AUTONOMOUS WEAPONS

Countries that are developing autonomous weapon capabilities are doing so as a result of their military's ambitions and the compelling reasons to adopt such technology. One of the obvious reasons for developing autonomous weapons is the reduction in risk to a military's soldiers while still retaining the ability to project power (Horowitz 2019: 769). Cost and vulnerability calculations have resulted in the US Navy pursing the development of Unmanned Underwater Vehicles (UUV) as well as Unmanned Surface Vessels (USV). It is plausible that a navy is able to acquire hundreds of robotic vessels for the price of a medium sized modern warship and, in so doing, increase its force mass. Modern warships are prime targets for the enemy and losing a few robotic ones will not likely impact the outcome of battle. The US Army is also thinking along the same lines in looking to replace humans in highly exposed frontline engagements (Klare 2019: 9). One of the more probable places where militaries are likely to deploy LAWS is in settings that are unambiguous and where civilians are not likely to be found, for example in either naval or air-to-air combat. The

necessity of being able to fight at machine speed is also likely to have a major impact in the advancement of automation in the military (Horowitz 2019: 770).

A further incentive to adopt LAWS is the role they would play as a force multiplier. This would result in a reduction of soldiers required for specific missions because the effectiveness of each soldier would be higher. In addition, LAWS would be well suited to tasks deemed dirty, dull, and dangerous. Dirty tasks involve situations in which soldiers are exposed to radiological or chemical material, dull tasks have a long duration, and dangerous tasks are best explained as bomb disposal (Etzioni & Etzioni 2017: 72). It would also enable the military to dramatically increase the number of missions it conducts simultaneously across multiple geographies with few humans involved in tactical execution of the mission.

Furthermore, as unmanned systems are tethered to their human operators (Schmitt & Thurnher 2013: 238), communication networks are needed to maintain contact with the remote controlled Unmanned Combat Ariel Vehicles (UCAV) and these networks are prone to interference through electronic countermeasures as well as kinetic attacks on the infrastructure that supports these networks. In highly asymmetrical conflicts, these communication links are generally not endangered; however, this could change if the conflict is with a more sophisticated foe (Altmann 2019: 3). LAWS, on the other hand, will still be able to continue with the mission, even if communication links have been severed (Haas & Fischer 2017: 289). Therefore, as highlighted by Altmann & Sauer (2017: 119), LAWS could eliminate the need for constant control and communication links. Currently, these communication links need to be daisy-chained, meaning there are a number of links in a communication network and no direct connection between the controller and the UCAV. Furthermore, each link in the communication network needs to remain in a line-of-sight with one another to remain connected. While these communication links already allow for communication with and control of UCAV without disclosing the location of the system, eliminating them all together would provide better assurances that the links will not be disrupted or hijacked. There would also be a tactical advantage in battle over a system that is remotely controlled as the delay between the operator input and response of the weapon would be non-existent in an autonomous system (Altmann & Sauer 2017: 119).

The compressed decision cycles achieved through machine speed warfare can convey a tactical advantage and, in turn, a strategic advantage which also adds pressure to the adoption of autonomy. Autonomous weapons can enable the ability to employ force in a more discriminate and effective way (Boulanin & Verbruggen 2017: 63). Autonomous weapons whose targets are derived through the algorithmic analysis of battlefield intelligence could also enable more accurate targeting and, in so doing, reduce collateral damage and public outrage (Haas & Fischer 2017: 289). Autonomy also provides the possibility for collaboration between weapons systems and can allow them to operate en mass in what has been dubbed swarming. This would allow greater strategic, structured, and coordinated collaboration than if controlled or commanded by human operators. The advances in the underlying technology of collaborative autonomy as well as low-cost robotics platforms have the potential to reintroduce mass which is what modern militaries no longer have, owing to the large cost of modern weapons systems (Boulanin & Verbruggen 2017: 63).

A further reason to adopt autonomous technology is that it does not have the same physical and mental limitations when compared to human soldiers. Autonomous weapons do not suffer from confusion, fatigue, and exceeding the limits of the human mind that often result in errors on the battlefield. For example, autonomous snipers would never hesitate after days of waiting in the same position. AI can also optimise tactical activities. AI would be able to outperform a seasoned battle commander at storming an enemy position by coordinating manoeuvres and fires across automated and networked platforms (Horowitz 2019: 770). The immense mental and physical strain fighter pilots endure through having to perform extreme g-force manoeuvres while maintaining high levels of concentration as well as situational awareness make them prone to exhaustion and fatigue and subsequently errors. Robot pilots would suffer no such maladies and, in fact, could perform even higher g-force manoeuvres and have quicker reaction times (Etzioni & Etzioni 2017: 73; Altmann & Sauer 2017: 123).

Finally, automation presents the potential to reduce the number of humans staffing militaries and the complexity involved in recruiting, training, and caring for human soldiers (Horowitz 2019: 770). It has been estimated that the size of a brigade can be

reduced by 25% without reducing its effectiveness by increasing the use of support robots (Etzioni & Etzioni 2017: 72). This presents real potential for cost savings for militaries. For example, combat aircraft pilots are expected to keep their skill set sharp and can only do this through continuous practice in real world environments. These flights and costs would not be needed in the case of automation as autonomous weapons do not need to practice continually to combat degradation of their operational effectiveness (Boulanin & Verbruggen 2017: 63).

There are clear advantages in adopting autonomous capabilities, from cost and causality reduction, to force multiplication, and the increase in the speed and tempo of warfare. On the surface it would almost seem as if not adopting autonomous technology would be a grave strategic error for the states that possess this technology and can harness it for such purposes.

2.8. DISADVANTAGES OF ADOPTING AUTONOMOUS WEAPONS

As Hoffman *et al*. (2016: 248) argue, the promise that LAWS will potentially bring faster, cheaper, and better military systems is nothing but unbridled enthusiasm from technologists. They point the fact that the US armed forces have been using automated weapons systems for two decades to protect its forces from attack and strike the enemy from long range. These examples include the Patriot missile defence system which can identify, track, and engage enemy warplanes and missiles without human approval. However, this automation has not reduced the proficiency or quantity of trained war fighters needed. In fact, in the case of the Patriot missile system, more highly skilled soldiers with greater experience were required and failures in training with regards to the system have led to the loss of innocent life through incidents of fratricide. LAWS would also degrade the control and predictability of events during combat. There is also the challenge that LAWS would be vulnerable to cyber attacks through direct hijacking of the weapons or the control system being hacked and tampered with (Altmann 2019: 4). At the moment the advantages and disadvantages of LAWS are theoretical as there is no state that employs a force that is comprised of only autonomous weapons. However, as highlighted above, we can draw some

conclusions for the autonomous systems already in place and attempt to project that going forward.

2.9. DO AUTONOMOUS WEAPONS ALREADY EXIST?

Weapons systems exist that, once deployed, are capable of independent target detection, identification, tracking, selection, and making the decision to engage that have been used for decades. These existing weapons can be grouped into the following categories: robotic sentry weapons, active protection systems, air defence systems, loitering munitions, and guided munitions (Boulanin & Verbruggen 2017: 36). For example, the Israeli Harpy and the Brimstone missile used by the United Kingdom (UK) would both qualify as an AWS (Horowitz 2016a: 91).

The other class of weapons that are already deployed that are capable of autonomous functions are missile defence and counter rocket systems like the US Navy's Aegis system and the land-based Patriot missile defence system, the Israeli Iron Dome counter rocket system, as well as the Phalanx Close-in-Weapons System (CIWS) used by the US Navy. However, these weapons are not intended to target humans and are designated anti-material weapons. While these weapons have autonomous features, and although they could be considered forerunners of LAWS, they are not LAWS (Mull 2018: 485-486) because the examples above are weapons systems that have a particular purpose and are generally fixed in their location, for example the air and projectile defence systems. These anti projectile systems are perfectly suited to be autonomous given the speed at which they need to identify, track, and engage threats. However, this is a far cry from a robot being capable of moving around on the battlefield and performing a wider range of autonomous tasks, including both offensive and defensive manoeuvres.

That said, in 2020 the Libyan Government of National Accord (GNA) attacked forces aligned with Khalifa Haftar know as Haftar Aligned Forces (HAF) using Turkish made STM Kargu-2 drones. These drones were used in a true autonomous mode with no communication links between the operator and the drone and they were used as a force multiplier that was used to slowly degrade the HAF operational capabilities. HAF

air defences were proving ineffective against the drones. Because HAF forces were not trained and didn't have the motivation to fight against this new technology, they retreated in disarray and were subsequently subjected to continual harassment by the drones (UNSC 2021: 17). This is the first publicly available documented use of LAWS in a combat situation.

2.10. WHAT HAPPENS IF AUTONOMOUS WEAPONS FAIL?

An additional challenge that Sharkey (2012: 790) mentions is that, if a machine fails, it can fail in way that no human would. While no example of autonomous weapons failing currently exists, we do have a few examples of the damage that can be done using autonomous systems in the financial world, including the "flash crash" which occurred on 6 May 2010. A flash crash was experienced on the Dow Jones Industrial Average (DJIA) which resulted in 10 percent of the value of the index being wiped out in a matter of a few minutes. The US Security and Exchange Commission (SEC) conducted an investigation and determined that a stock trade that was automated by a sell algorithm had executed a large sale order which was supposed to be broken up over a number of hours in less than 20 minutes. The result of this sell algorithm interacting with other high frequency trading algorithms was a very fast price drop. This is an example of an unintended interaction with other automated systems that resulted in the amplification of the initial effect as it starts a cycle of a large order which lowers the price. This triggers other algorithms to sell which, in turn, lowers the price further (Scharre 2016: 36). In another example, Knight Capital launched an automated trading algorithm that seemed to be trading by selling low and buying high many times per second, losing roughly between $10 and $15 a trade. This carried on for 45 minutes and resulted in a loss of more than $440 million before it was shut down (Harford 2012). While these are not autonomous weapons, there is a cautionary tale about the interaction of automated algorithms and the resulting consequences when things go wrong.

2.11. A NEW ARMS RACE

The science and technology underpinning LAWS will, through the development and application to everyday life such as the development of autonomous vehicles, be a very general technology. The everyday use of this technology would make controlling the development of these weapons very hard and whether this develops into an arms race very much depends on the political undercurrents underpinning the approach to LAWS (Horowitz 2019: 766). With this in mind, world powers have added to the narrative. In 2017 Russian president Vladimir Putin said that the nation that ends up becoming the world leader in AI will rule the world, in 2017 China stated that they wish to become the world leader of AI by 2030, and the White House released the American AI Initiative and the US DoD published an AI strategy in 2018 (Scharre 2019: 135).

Wong *et al*. (2020: 61) raised the concern that the development of LAWS might result in arms race instability, with an arms race potentially imminent amongst China and the US. They also argued that the proliferation of LAWS would lead to a classic case of a security dilemma as a result of a push for countermeasures which would exacerbate uncertainties, leaving states feeling less secure. However, this developing narrative about an “AI arms race” highlights a misguided understanding of the nature of the risks of AI the as well as the additional novel risks that it poses. In fact, the actual risk here is not that a country will lag behind its peers in the development of AI technology, but rather that the perception of a race may result in countries cutting corners in a rush to deploy these weapons. This will result in a proliferation of unsafe AI systems which would endanger the countries that deployed them as much as their adversaries (Scharre 2019: 135). This is an important point as the mere perception of an arms race could have dire consequences in the proliferation of unsafe LAWS.

As Horowitz (2019: 776-777) suggests, it is crucial to differentiate between the proliferation of a military technology and an arms race. Huntington (1958: 41-42) defines an arms race as a progressive and competitive acquisition of arms by either two states or a coalition of states as a result of incompatible objectives or mutual fear. This results in a form of reciprocal interaction between the two actors. Therefore, a state can increase its acquisition of arms either through perceived necessity or economic necessity independently of the actions of other states. Many arms build-ups

have been incorrectly classified as arms races as these increases have been without reference to a particular conflicting difference between states.

Therefore, the debate about a LAWS arms race is actually more realistically a proliferation debate and some amount of proliferation is unavoidable largely due to the underlying factors necessary to produce LAWS (Horowitz 2019: 777). LAWS do not need a military specific development path similar to that taken for nuclear weapons. As AI and robotics technology matures and permeates everyday civilian life, militaries will be able to adapt the technology to suit their own needs. Therefore, the development of LAWS will benefit from the adoption and customisation of civilian technologies (Altmann & Sauer 2017: 124). According to Horowitz (2019: 777), military technology diffuses more rapidly when there are underlying commercial demands and applications beyond the scope of military use and the per unit cost is low. An example is the significant reductions in price and size and an increase in durability of Light Detection and Ranging (LIDAR) systems used in autonomous cars to give a 360-degree image of their environment (Altmann & Sauer 2017: 124-125).

Commercial demand and use of AI is growing across the world. In fact, AI developed for purely commercial use could have a spillover impact on the development of military technology. This would reverse the previous trend of military technology spilling over into the civilian realm. One such example is the development of the Global Positioning System (GPS) (Horowitz 2019: 777-778). Now that LIDAR units are commercially available and affordable and are crucial in giving a mobile system autonomy, there is no reason not to imagine that militaries will adopt, adapt, and refine the technology for their own needs (Altmann & Sauer 2017: 124-125). As mentioned, AI is an enabler technology with broad military and civilian applications, in contrast to stealth technology which ultimately only has military applications. This broad applicability and commercial availability provide the avenue for less wealthy states and non-state actors to modify the current available technology to build rudimentary LAWS (Horowitz 2019: 777-778). This increases the danger that, even if the world moves to limit the use of autonomous weapons, terrorist groups could devise weapons with store-bought autonomous functionality to spread fear in civilian populations.

This also changes the narrative of who is driving the Research and Development (R&D) of revolutionary military technology. In the past, the development of nuclear weapons or new conventional weapons systems such as the F-35 multirole stealth fight jet was driven by the military industrial complex. Now, most of the revolutionary components for LAWS will likely be driven by universities and commercial enterprises that are subject to market forces which will drive costs down and shorten innovation cycles. Therefore, LAWS will be revolutionary on the battlefield but evolutionary in technological development making use of commercially available technology (Altmann & Sauer 2017: 125). This change in approach radically shifts the idea of who is driving the development of LAWS and will ultimately make any attempts to meaningfully regulate or control the diffusion of this technology almost impossible.

Arms races are traditionally a concern for the world's super powers because the attempt to gain a competitive edge or to prevent falling behind often leads to excessive and destabilising accumulation of weapons. A race for AI-powered weapons could prove dangerous because the consequences and untested nature of these weapons are unknown and have the potential to have a catastrophic impact. In fact, this could lead to a scenario where weapons are deployed without fully understanding their capabilities and limitations which could result in unintended fatalities or uncontrollable escalation (Klare 2019: 8). As this is a new technology, unknown factors exist as to what a race to gain the upper hand in AI could result in and what tensions this could generate in the international arena.

2.12 FUTURE WARFARE: DRONE SWARMS

Distributed collaborative systems also known as drone swarms are groups of UCAVs that can move and act together with very little to no human interaction. The US Navy proved that they can deploy drones mid-flight from F/A-18 Super Hornets over California. These drones flew in formation and exhibited the ability to make collective decisions in what was the largest micro drone swarm in the world at the time (Lachow 2017: 96). With the use of 3-D printing, these drones can be produced in large batches and can be deployed from a range of other flying platforms (Altmann & Sauer 2017:123).

Swarm behaviour can be observed in nature in schools of fish or flocks of birds which rapidly change direction in unison in what would appear to be highly choreographed movements. Swarming systems are made up of individual agents which interact with each other and the surrounding environment. With the individual agents following simple rules, the collective outcome can result in sophisticated and complicated behaviour, including emergent intelligence (Lachow 2017: 96). Swarming is a key capability and can enable saturation attacks which is what will be required in order to overcome existing extensive range defensive capabilities, including those capabilities still in development (Altmann & Sauer 2017: 127).

Autonomous weapons, including drone swarms, are already impacting the international security environment. Drone technology is being actively developed by a number of nations, including China, Russia, and South Korea. China has tested a swarm of over 100 fixed wing drones and over 1000 roto-wing drones.

Furthermore, drone swarming technology has the ability to significantly impact chemical, biological, radiological, and nuclear (CBRN) weapons and their related defences. The first of these impacts is the ability of drone swarms to locate dispersed mobile missile launchers (Kallenborn & Bleek 2018: 525).

Another aspect of this phenomenon is that of human machine teaming, Centaur Warfare, or manned-unmanned teaming (Haas & Fischer 2017: 288; The Economist 2020). This will allow humans and machines to work together through a redistribution of the cognitive workload where the soldier would be able to hand some tasks over to a computer while managing the remaining tasks. An example is Lockheed Martin's missile avoidance system which is capable of determining which plane in a formation is the target of an attack and what evasive manoeuvres are required to avoid it. In contrast, a human being would be required to interpret the same using a number of displays. As mentioned above, the Auto-GCAS ground collision avoidance system prevents planes from colliding with the terrain if a pilot is distracted, disoriented, or unable to respond (The Economist 2020).

2.13. FUTURE WARFARE: LAWS AND THEIR POTENTIAL IMPACT ON NUCLEAR SECURITY

Johnson (2020: 26) has argued that not only would drone swarms be well-suited to Intelligence, Surveillance, and Reconnaissance (ISR) missions as suggested above, they could also possess a first strike capability against an enemy's mobile nuclear capabilities, including both mobile missile launchers and possibly submarines. Some have argued that autonomous surface vehicles like the DoD's Sea Hunter might erode confidence in a nation's second-strike capability by making deterrence submarines redundant (in making the depths of the sea transparent) and that the proliferation of AWS could have major strategic consequences for nuclear security and the possible rapid escalation of future conflicts (Johnson 2020: 30).

2.14. THRESHOLD FOR CONFLICT

It is possible that LAWS could lower the threshold for conflict by disturbing the deterrence balance between adversaries and encouraging offensive use. Furthermore, by reducing human soldiers involved in a conflict, the body bag count of said conflict can be reduced, which could incentivise starting wars. These positions focus on the assumed asymmetry in the military capability of advanced militaries when deploying LAWS and are based on the premise that war is not morally acceptable, and that the development and deployment of LAWS would increase conflicts. However, as compelling as these arguments may be, there are a number of serious disincentives to start a war. These include military expenditures, international and domestic pressure, and international law (Petrenko 2012: 84-85).

To support this argument, an analysis of war gaming by Wong *et al*. (2020: 60) found that LAWS may enhance the credibility of the US's extended deterrence capability by lowering the risk to US military personnel when using LAWS. The ability of LAWS to remain strike-ready for extended periods of time, faster decision cycles, and potentially greater precision than humans, could lead enemies to conclude that the US will be more willing to deploy these technologies when faced with threats to their allies. However, this can also lead to a reduction in the US's ability to assure their allies that

they would put US lives on the line in times of crisis. Widespread use of LAWS could increase the risk of conflict escalation and crisis instability due to tasks being executed at machine speed with inadvertent escalation being a real concern. Concerns have also been raised about the possibility of first strike instability in order to gain a military advantage or prevent an adversary from gaining a military advantage (Wong *et al*. 2020: 60-61).

2.15. CONCLUSION

This chapter provided an overview of the concept of AWS, the underlying technology that makes these weapons systems possible, the current state of development of these weapons, and their current and future uses. The future is less about combating the arrival of science fiction-style terminator robot assassins but rather the commercial development of AI technology that, if adapted for military purposes, could lead to weapons systems capable of acting on their own. AWS are already used in a number of applications by militaries across the globe, but none that are capable of fully autonomous action outside of their narrow scope of function. The danger also exists that commercially available technology can be adapted by terrorist groups for their own purposes. It was also highlighted that there are numerous incentives to adopt autonomy in military capabilities; however, the security dilemma this presents could result in a technology arms race to secure the leading position in AI-based weapons technology.

3. CHAPTER THREE: LEGAL CHALLENGES TO AUTONOMOUS WEAPONS

3.1. INTRODUCTION

The development of autonomy in weapons systems will directly challenge the current legal understanding of what a weapon is as well as the means and methods of warfare that are used. The incorporation of AI and the resulting possibility of autonomy in weapons systems place LAWS in a distinct category, separate from all forms of prior weapons systems. All weapons throughout human history have been passive implements and inert tools (Liu 2012: 628) used and directed by humans and were not a force unto themselves (Malik 2018: 621). Humans have traditionally controlled and manipulated these implements to inflict damage, cause injury, and act violently towards other humans (Liu 2012: 628). With the arrival of LAWS and, to a smaller degree Remote Weapons Systems (RWS), the use of force may require limited human decision-making or oversight (Liu 2012: 629). This means humans could foreseeably relinquish meaningful control and oversight and, in effect, outsource life and death decision-making to machines that have the potential to behave in unforeseen ways in the heat of battle (Malik 2018: 621).

LAWS seem to inhabit a grey area between existing legal categories of being both a combatant and a weapon. Classifying LAWS as a weapon fails to acknowledge that they can directly commit violent acts and does not take into account the levels of autonomy that they can possess. On the other hand, when LAWS are classified as a combatant, the protections afforded to human combatants in law infer the omission of machines. This indicates both practical and conceptual hurdles that would prevent LAWS from being classified as combatants but assigning LAWS to the category of only weapons would, at best, only partially consider the true impact of these weapons systems and the real challenge they pose (Liu 2012: 629). Therefore, LAWS would be unique in that they would change how we think about who conducts warfare and how it is conducted. This results in the need to evaluate the challenges LAWS would poses to the existing legal framework.

3.2. JUST WAR THEORY

Just War Theory is a doctrine that regulates both the reasons for going to war, referred to as *jus ad bellum*, as well as how the practice of war should be conducted, referred to as *jus in bello*. The traditional view of Just War Theory states that there are three fundamental principles that control how war should be conducted. Those are military necessity, distinction, and proportionality. The task of these three principles is to minimise the violence and damage committed against non-combatants and their property. If harm were to occur to these groups, even if it was foreseeable, it must be of an unintended or collateral nature (Guersenzvaig 2018: 58). These principles will be further discussed below.

3.3. INTERNATIONAL HUMANITARIAN LAW

International humanitarian law is the term that describes the current interpretation of *jus in bello* – which are the laws governing how warfare is conducted (Alexander 2015: 110). Put another way, IHL is the operationalisation of the abstract principles contained in *jus in bello* through treaties and conventions contained in the four Geneva Conventions of 1949 as well as the Additional Protocols of 1977 (Guersenzvaig 2018: 58). The Additional Protocols I (AP I) and II (AP II) are supplemental international treaties for the Geneva Conventions of 1949 which extend the legal protection of the wounded and civilians. For the first time, it also established detailed humanitarian rules covering civil wars (ICRC 2009). IHL has evolved over more than a century and is principally concerned with two aspects balancing humanitarian wellbeing and military necessity. The first aspect it is concerned with is protecting civilians from conflict and preventing soldiers from being subjected to unnecessary cruelty and suffering, while the second aspect it is concerned with is allowing military objectives to be attained (Wagner 2014: 1384). The ICRC has a special relationship with IHL and acts as its promoter and guardian thereof (Alexander 2015: 110). The ICRC describes it as follows:

> *International humanitarian law is part of the body of international law that governs relations between States. IHL aims to limit the effects of armed*

> *conflicts for humanitarian reasons. It aims to protect persons who are not or are no longer taking part in hostilities, the sick and wounded, prisoners and civilians, and to define the rights and obligations of the parties to a conflict in the conduct of hostilities* (ICRC 2010).

Generally speaking, IHL is the division of public international law that moderates the conduct of armed conflict and the suffering it produces. It is the legal framework which is concerned with protecting civilian populations and treating soldiers humanely while allowing militaries to reach their objectives. IHL used to refer to the Geneva part of *jus in bello*, as it contained a humanitarian emphasis, whereas the Hague law focused mostly on methods of warfare. This differentiation is artificial and has been called two sides of the same coin as they are both rooted in humanitarian considerations. Thus, IHL can be used to reference all the facets of international law that are concerned with armed conflict, whether that be conventional, customary, Geneva, or the Hague law (Alexander 2015: 111).

3.4. LEGAL CHALLENGES: PROPORTIONALITY AND DISTINCTION

It can be argued that no issue emphasises the challenge that LAWS pose more than the application of existing legal frameworks and the need to guarantee the adherence to the cardinal principle of distinction and the resulting principle of proportionality (Beard 2014: 663). The principles of IHL endeavour to give protection to civilian populations but there is no way to eliminate civilian injury and death from conflict (Foy 2014: 55). As part of the century long evolution of IHL the two principles of proportionality and distinction have been developed and represent the tension between the disparate goals of reducing suffering and advancing military ends (Wagner 2014: 1384).

AP I entered into force in 1978 (Wagner 2014: 1385) and at present has 174 states as parties (ICRC 2021) which represents a significant proportion of countries in the world. It also includes many states that are likely to pursue the development of LAWS, including Russia, China, Germany, the UK, and France. However, some states, including the US, India, and Israel are not parties to AP I. It was conceived by states

in order to make IHL not only more universal and complete, but also to update it to be applicable to modern conflicts (ICRC 2009).

While the definition of the principles of both proportionality and distinction are not explicitly contained in AP I, both concepts are included and are well established in customary international law which makes their explicit absences from AP I less significant. In fact, the US has made consistent statements upholding the customary nature of most of the provisions of AP I and has incorporated either similar or identical language used in AP I in their military manuals, including the principles of proportionality and distinction (Wagner 2014: 1385) However, with the current technological state of the development of autonomous systems, these weapons will struggle to implement both the principles of distinction and proportionality as will be discussed below.

3.4.1 The Principle of Distinction

Just War Theory calls for the principle of distinction to be observed during conflict. This requires that military engagement must differentiate between both civilians and combatants, including the differentiation between civilian and military objects. This differentiation between the civilian population and their objects and military personnel and their objects is the first stage in determining if someone or something can be lawfully targeted. IHL assumes that if a person is not participating in hostilities, they are a civilian. This idea was integrated into the 1868 Declaration of St. Petersburg which is the first mechanism in IHL. This rule has been further reinforced in the annex to the Hague Convention Respecting the Law and Customs of War on Land as well as the AP of the 1949 Geneva Conventions (Wagner 2014: 1388-1389). Article 48 of AP I states the following:

> *In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives* (ICRC 1977).

AP 1 would still apply to states that are not signatories as the principle would still be applicable through customary international law (Foy 2014: 54). The subsequent provisions go further to prevent individual civilians from being targeted unless they are directly taking part in the hostilities. It also outlaws the targeting of places of worship, works of art, or historic monuments and prohibits the attack on the natural environment and objects that are deemed necessary for the survival of the civilian population – installations that, if attacked, could do great damage, for example, dams or nuclear power stations (Wagner 2014: 1389). Indiscriminate attacks are also prohibited and, central to this, doubt. If the attacker has some doubt, or if another reasonable combatant placed in the same situation would hesitate, the attack should not be carried out. This prohibition is contained in Article 50 of AP I and is also accepted as customary international law.

Applying the principle of distinction has become more difficult as conflict has evolved from occurring between state parties to counterinsurgency operations. However, this has no impact on the obligation to distinguish between military and civilian targets (Foy 2014: 54-55). Distinction is concerned with the identification and separation of targets between civilian and military and the exclusion of all civilian targets if they are not directly participating or aiding hostilities. It even goes as far as to prohibit strikes that target military assets if it could harm the civilian population. It also introduces the requirement of certainty when committing to the attack to ensure that civilians and their infrastructure is not damaged unnecessarily.

3.4.2. The Challenge LAWS Pose to Adhering to Distinction

Advocates for the advancement of autonomy in weapons systems argue that, in the long run, LAWS could be more effective at applying the principle of distinction and thus be ethically more preferable. There is also the challenge of classifying a weapons system as inherently indiscriminate when it has been specifically designed to attack military targets. However, autonomy does raise a question that is not directly addressed within in the current IHL framework about the level or quality of human judgement that will need to be exercised when applying both the principles of

distinction and proportionality. The need to consistently distinguish between civilians and combatants is an enormous challenge scientists and weapon designers building LAWS currently face (Beard 2014: 664).

As Sharkey (2012: 788) highlights, LAWS would not be capable of discriminating in any form in order to meet the requirements of the principle of distinction. Currently, LAWS are not capable of distinguishing between combatants; civilians; or other immune actors such as medical workers, retirees, and soldiers who have surrendered, or are wounded or mentally ill. Current autonomous systems such as the Israeli Harpy could be classified as having a weak ability to discriminate. It is an anti-radiation munition that scans for radar signatures. Once detected, it matches these radar signatures to a database to determine friend or foe status. If the radar signature registers as a foe, the Harpy will dive bomb it. While this is a type of distinction, it does not satisfy the IHL principle if distinction because the Harpy would not be able to discern if the radar is located in on a remote anti-aircraft station or perhaps located on top of hospital.

Sharkey (2012: 788-789) goes on to highlight the three challenges of adhering to the principle of distinction. The first is that a computer lacks the sensory capacity to determine whether or not a person is a combatant or a civilian, especially when considering insurgent warfare or wounded or surrendering combatants. Using current sensory technology, it is possible to identify whether something is human but not much more. This is further complicated by moving targets or by having to identify them either from the air or in the fog of war.

Secondly, computers are capable of computing any procedure that can be transcribed into programming language. In other words, comparable to a formula or recipe, all the elements need to be specified in sufficient detail for the computer to execute it. The challenge with the principle of distinction in this situation is that a definition of exactly what constitutes a civilian does not exist. Therefore, not enough detail exists to trust that the computer will execute the task correctly all the time.

Finally, even if the sensory equipment was able to distinguish the difference between a civilian and a combatant, key information such as contextual and situational

awareness, and even common sense that is crucial in discrimination decisions, would still be missing. This might be possible in the future in some narrow applications but the ability to discriminate on a human level or having some ability to apply common sense reasoning might remain computationally impossible.

This is a critical legal challenge to the development of LAWS as the current technology does not allow for the principle of distinction to be adhered to all the time with certainty when considering an operational LAWS.

3.4.3. The Principle of Proportionality

Article 51(5)(b) of AP I states that "an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated" (ICRC 2005a). Therefore, the article requires that any attack which could cause harm to civilians have an adequate military objective without inflicting excessive civilian harm (Wagner 2014: 1385). Thus, the target considered in the attack must be proportional (value equal to) to the potential collateral damage, including loss of civilian lives. While there is no reference to proportionality in AP II which applies to the law of intranational armed conflict, it is considered as an inherent principle of humanity and, as it is included in the preamble of AP II, it must be included in the application thereof. In addition, the ICRC found no contrary practice to the principle of proportionality and has argued that it has crystallised into customary law (Foy 2014: 55).

The concept of military advantage has been interpreted to include the advantage that is projected to occur as a result of an attack as well as the projected advantage to the military operation in its entirety. The military advantage must also be tangible and direct, which is interpreted to mean that the advantage must be substantial and must be applicable in the near future while small advantages or those that will only be evident in the long run must be ignored. For example, in Canada's Law of Armed Conflict this understood to indicate that a commander must have a sensible and truthful expectation that the attack under consideration would make an appropriate

impact to the achievement of the overall mission. Therefore, proportionality demands a contextual balancing between the military advantage that might be obtained as a result of an attack and the potential injury to civilians and their property. Currently, the objective evaluation of the potential for civilian harm is possible and commanders often use collateral damage simulators in an attempt to ensure that their attacks are proportional. However, the determination of military advantage is more discretionary and contextual and necessitates asking whether a reasonable commander would come to the same conclusion (Foy 2014: 56).

3.4.4. The Challenge LAWS Pose to Adhering to Proportionality

For the principle of proportionality to be adhered to, a calculation would need to be performed by LAWS when an attack is being considered. It would have to evaluate the military advantage in comparison to the civilian lives lost or property that will be damaged and then decide if it is appropriate to attack. While you could potentially model the collateral damage there might be a challenge in trying to model the military advantage as well as the reasonable commander test given the contextual and subjective nature of those elements.

As such, the proportionality principle is often indicated as a major obstacle that would prevent LAWS from being compliant with IHL. It is important to investigate whether the ability of LAWS to apply proportionality in combat would impact the legality of the weapons and thus their use (Boogaard 2015: 247-250). While Arkin (2009: 33) argues that LAWS would be better at proportionality decisions, Sharkey (2012: 789) states that LAWS would lack the situational awareness and the agency to make proportionality decisions. He believes that this is in context of what is described as the easy proportionality problem, which is the reduction of collateral damage through the choice of the most appropriate weapon or munition and tailoring the use of it to the situation at hand. He points out that a computer programme called Bugsplat already exists and is in use by the US military for this exact purpose. This approach, however, can only minimise collateral damage, it cannot eliminate it. For example, if there is a target near a school with 200 children, the programme may recommend a course of action and choice of munition where only 30 children are killed instead of all 200.

Conversely, the difficult proportionality problem is choosing to use kinetic or deadly force within a given situation in the first place (Sharkey 2012: 789). The challenge here is that, in applying proportionality, there are inherently complex and subjective decisions that need to be made in order to prevent excess civilian casualties and property damage (Beard 2014: 664). When determining military advantage, there is allowance for operational discretion. It is in the contextual nature of conflict and the application of discretion in determining the military advantage to be gained that LAWS would be incapable of adhering to the principle of proportionality (Foy 2014: 56). The use of force in proportion to the military advantage gained is simultaneously a subjective and a qualitative decision and, as such, should be made by situationally aware, experienced, responsible, and accountable commanders who are qualified to make these decisions (Sharkey 2012: 789-790). These decisions often test the bounds of even the most rational and experienced commander and would most certainly defy resolution though a rules-based approach or even the most advanced AI system at present (Beard 2014: 664).

As Sharkey (2012: 790) states, war is more of an art than a science and the balancing tests encapsulated in the proportionality principle require human judgment with regards to the value of human life which must be considered in both the context of instantaneous facts presented on the ground as well as the more obscure long-term goals of engagement (Beard 2014: 665). Therefore, LAWS would be required to understand, interpret, and act on these nuances when making battlefield decisions regarding proportionality that inherently comes from training, experience, and just from being human. It is that nuance that would be very hard to capture in AI systems.

3.4.5. Could LAWS be Declared Illegal: The Case of Nuclear Weapons and IHL

In the case of nuclear weapons, the International Court of Justice (ICJ) issued an advisory opinion addressing the idea that nuclear weapons were not illegal *per se* because they were not necessarily intrinsically incapable of proportionality or distinction nor would they cause unnecessary suffering or superfluous harm in all circumstances. This would lead to the assumption that, if nuclear weapons are not

unlawful *per se*, then it is very unlikely LAWS would be declared illegal (Foy 2014: 53). Furthermore, the court emphasised the significance of the principle of distinction as well as the protection of civilians and emphasised that civilians must never be the target of an attack. These were singled out as the cardinal principles which constitute the fabric of IHL (Szpak 2019: 122). The principle of distinction was raised by the court to the level of *jus cogens*[3] when it stated that it is an intransgressible principle of international customary law (Wagner 2014: 1390). Therefore, using this line of argument, LAWS are unlikely to be deemed illegal. However, they will critically need to be able to distinguish if they are to be considered for potential legal use.

3.4.6 The Martens Clause

Righetti (*et al.* 2018: 125) suggest that if a weapon complies with both proportionality and distinction, it by no means automatically makes it legal. Ethical aspects should also be considered. However, ethics are not explicitly outlined in any of the relevant law. The Martens Clause, which is also known as the dictate of public conscience, states the following

> *Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity, and the requirements of the public conscience* (ICRC 1899).

This has been interpreted to mean that just because a weapon is not explicitly outlawed, does not mean that it can be used as there are ethical aspects that also need to be accounted for. In this case, societal and ethical considerations, expectations, and acceptance need to be considered. It could be argued that public conscience played a pivotal role in banning blinding lasers, chemical weapons, and APL (Righetti *et al.* 2018: 125).

---

[3] *Jus cogens* refers to the idea that certain principles shape the norms of international law and that these norms and resulting law cannot be derogated (Oxford Bibliographies 2015).

3.5. BEYOND LEGALITY: THE RESPONSIBILITY AND ACCOUNTABILITY GAP

In addition to the evaluation of the legality of LAWS, there is also the challenge that LAWS pose to the enforcement of international law. A specific importance has been placed on incorporating concepts like responsibility, accountability, and liability into legal codes during the codification of laws to allow natural persons and legal instruments to charge legal entities like governments and private companies for damages inflicted during certain events. In addition, significant attention has been paid to the establishment of strong institutions and their associated enforcement mechanisms in order to punish those responsible for crimes committed, especially in international law.

Although liability and responsibility are distinctive concepts used in law, accountability – which is sometimes used as a synonym for responsibility – does not have a strict legal definition, notwithstanding it being used heavily in the practice of international law (Malik 2018: 616). The ability to assign responsibility is crucial as it is a pre-condition for conducting a just war. It is important to be able to hold an individual responsible for the death of enemy soldiers and even more so for civilian deaths caused during the course of a conflict. This responsibility extends beyond causal responsibility and into the domain of a legal and moral responsibility and can be assumed to be one of the conditions for *jus in bello.* If we are to ever govern war with any semblance of morality, enemy soldiers should be offered the minimum sign of respect in that someone should be held responsible for the choice that was made to end a soldier's life. If this is not done, enemy soldiers will be treated the same way we treat vermin and can be killed with no moral regard.

Practically, if a weapon or means and method of warfare make it impossible to identify and subsequently hold someone accountable, then it's in conflict with the requirements of *jus in bello* and, as a result, it would be unethical to apply it in warfare. The immorality of WMD and other weapons such as anti-personal mines are that they violate this condition because no one is held responsible for the choice of deciding who lives or who dies when they are used (Sparrow 2007: 68). This demonstrates the difficulty concerning LAWS: killing someone comes with legal and moral responsibility but there is no realistic prospect of an AWS bearing responsibility in practice.

Human accountability is challenged by LAWS and it is this accountability which is core to current IHL and the international criminal justice system which has developed in order to support it. LAWS challenge the normative framework of IHL and their development might be incentivised in order to present a situation where individuals and states can avoid being held liable for their conduct in war (Grut 2013: 14). Therefore, it is critical to ensure that individual accountability is ensured in order to avoid international law becoming nothing more than a *brutum fulmen* – a harmless thunderbolt or ineffective threat. The purpose of International law is not only to set the standards for what states, non-state actors, and their agents should adhere to but also to recommend the required penalties should they fail to adhere to those standards. IHL norms will be meaningless without the ability to assign accountability for failing to adhere to them and the preservation of global peace and security would be threatened if there was no accountability for violations.

Accountability is also key to the victim's right of remedy (Chengeta 2016: 5). A victim's remedy is comprised of three components. The first is that of access to justice which deals with a state's responsibility to remedy its victims. The second is that of the right to reparation which deals with the responsibility to prosecute the offenders as well as the responsibility of individual actors, corporate actors, and non-state actors who possess a duty to pay reparations if they are convicted. The last component is the right to access information which concerns the right to be presented the truth concerning the violation of the victim's rights (Chengeta 2016: 7). Currently, IHL relies on assigning accountability to individuals but, as LAWS are not people, the ability to assign accountability, and in so doing offer remedy to the victims, is non-existent.

It would be difficult to determine who should be held accountable for accidents if LAWS were used without the ability to discriminate. It has been suggested that the battlefield commander be held accountable as it was on their orders that something went wrong. However, this may not be a fair consideration given the commander did not program the weapon. Is it the manufacturer's fault or perhaps that of the policy makers who decided to deploy the weapons? Could it have been tampered with or even hacked by a third party (Sharkey 2012: 791)? With LAWS the notion of accountability becomes opaque as it can't be assumed that the system designers, software developers,

maintenance engineers, or the commander on the battlefield acted wilfully with the intention of committing the crime or that they are even aware of the exact circumstances in which the crime took place. Without intention there is no crime so who should be held responsible (Guersenzvaig 2018: 60)?

This is the accountability gap that is created with LAWS. Accountability is not only critical to international law; the victims' right to legal remedy is threatened when such an accountability gap is created (Chengeta 2016: 2). Assigning individual responsibility through the mechanisms of accountability acts as an exercise in retribution for the victims, while at the same time deterring others from acting similarly in the future. It is a necessary mechanism in fighting a just war (Guersenzvaig 2018: 60). Without the ability to assign direct accountability for the execution of atrocities by a badly programmed or out of control robot in order to enforce compliance with IHL, civilian lives and property can be placed in danger (Sharkey 2012: 791). With this understanding, it would be problematic to deploy LAWS in conflict situations.

However, no matter how much the autonomy of machines increases, they do not currently and will not in the future have moral agency. Machines would have to have the ability to think and exercise free will before they attain moral agency and, in so doing, assume legal responsibility (Chengeta 2016: 11). Using this reasoning, it is unlikely that LAWS will ever have any moral or other form of agency and therefore be held accountable for their actions (Sharkey 2012: 791).

Sparrow (2007: 73) makes the argument that the use of LAWS is comparable to the use of child soldiers when trying to assign wartime accountability. One of the unethical aspects of deploying child soldiers is that they cannot be held accountable for their actions after making life and death decisions. While child soldiers do not have full moral agency and, as a consequence are not able to be held accountable, they are autonomous at least to a greater extent than a robot would be, despite being incapable of comprehending the full moral dimension of their actions and being unable to connect their crime with their punishment. If one considers that those who deploy them do not control them, meaning they have limited, yet adequate autonomy, the problem with attributing the responsibility for their actions becomes clear.

The issue of child soldiers points to the possibility of armies that have no moral responsibility being capable of intelligent action. Child soldiers, and potentially LAWS, occupy a conceptual space where they are autonomous to the extent that attributing their responsibility to an appropriate adult would be difficult but not autonomous enough to bear responsibility for their actions. The idea of LAWS occupies this difficult realm for the foreseeable future and makes it difficult to hold someone responsible for taking a life unjustly. This comparison highlights the challenge with regards to agents that possess some autonomy while simultaneously being unable to hold them accountable for the crimes they commit.

Traditionally, one of the ways to ensure personal accountability was charging commanders with ensuring their soldiers adhere to the laws of war, as per AP I:

> *In fact, the role of commanders is decisive. Whether they are concerned with the theatre of military operations, occupied territories or places of internment, the necessary measures for the proper application of the Conventions and the Protocol must be taken at the level of the troops, so that a fatal gap between the undertakings entered into by Parties to the conflict and the conduct of individuals is avoided. (1) At this level, everything depends on commanders, and without their conscientious supervision, general legal requirements are unlikely to be effective. (2) Undoubtedly the development of a battle may not permit a commander to exercise control over his troops all the time; but in this case he must impose discipline to a sufficient degree* (ICRC 1987b).

With the power that they wield over their subordinates, commanders have the ability to guarantee compliance to the law by upholding discipline, being knowledgeable about the behaviour of their troops, and being in the position to take action to either prevent or punish any transgressions. Commanders are seen as being responsible for avoiding what was termed in the previous passage as the fatal gap between international legal obligations and the perpetrating of gross violations. Commanders have the authority to stop war crimes from happening and, in the case where they were not able to stop them from being committed, they have the power to punish those who transgressed to discourage others from transgressing in the future. Commanders who decide not to exercise this power can be held responsible for crimes that could

have been prevented as well as for crimes they did not punish (McFarland & McCormack 2014: 365). The basis for this was established when the Nuremberg Tribunal declared that "[c]rimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced" (International Military Tribunal Nuremberg 1946). Therefore, enforcement of *jus in bello* relies upon the identification and prosecution of those people responsible for gross violations of international law. LAWS would not qualify as an abstract entity, but the principle of the judgement is still applicable. Regardless of its sophistication, LAWS would be considered inanimate objects and would be no more responsible for perpetrating the crime as a rifle would be today (McFarland & McCormack 2014: 366).

Those in favour of LAWS argue that humans also commit atrocities in war, either on purpose or by mistake, so why would LAWS committing an atrocity by accident or on purpose be any different? Of course, atrocities that are committed on purpose are designated war crimes and those responsible are held accountable under IHL. The question then is what happens when LAWS commit a war crime? When discussing war crimes, the challenge is that, without intent, there can be no war crime. Can LAWS act with intent? When dealing with accountability in this regard, especially where no human is involved, the process of acting with intent to commit a crime needs to be looked at with the idea that LAWS will, at some point in the future, be deployed (Guersenzvaig 2018: 60).

Given that LAWS will be capable of destructive acts, that they will potentially be unpredictable, and that they will inevitably be deployed at some point, they are likely to be involved in an incident that has destructive and deadly consequences. If the assumption is made that no one in the incident acted irresponsibly or with intent, then under current international criminal law it will not be possible to hold anyone liable. The allocation of criminal liability to individuals for war crimes stems from the aspiration to hold people accountable for the crimes they commit and to discourage reoccurrences (Crootof 2016: 1350-1351). By definition, a war crime is a blatant contravention of IHL and would result in individual criminal liability. Criminal liability stems from the requirement that the individual concerned acted wilfully, which is interpreted as acting either recklessly or intentionally. LAWS threaten to destabilise

almost seven decades worth of effort to create a system of international criminal law by voiding the presumption that gross violations of IHL will not occur without wilful human action (Crootof 2016: 1350-1351). Therefore, LAWS present the possibility of having a weapons system that acts independently and without legal consequence under the current legal framework, necessitating a change in legal thought around responsibility and accountability in conceptualising how to deal with AWS.

3.6. TORT LAW AS A POTENTIAL SOLUTION

While several domestic legal themes have an international equivalent, for example intellectual property law and international intellectual property law as well as civil rights law and IHL, there is no international equivalent for tort law. Tort law is the legal framework that deals with non-contractual civil violations for which individuals are able to pursue remedy. Lawyers dealing with environmental issues have been attempting to establish an internationally applicable liability via tort law for multinational environmental degradation but these attempts have not yielded much. Article 75 of the Rome Statue of the International Criminal Court (ICC) states that the court has the ability to make a compensation order directly against an individual who is convicted and has the power to specify reparations. In the history of international law, this was the first occurrence where victims were able to individually pursue remedy from an international tribunal. By utilising tort law, one could potentially close the responsibility and accountability gap ensuring there are consequences for transgressions of international law by LAWS. However, the ICC does not seem willing to use Article 75 to warrant tort-like reparations (Crootof 2016: 1351).

The US uses domestic legislation called the US Alien Tort Statute (ATS) which is unusual in nature due to the fact that it recognises and creates federal subject matter jurisdiction in order to prosecute individuals for international torts. A proposal to enhance the ATS in order to allow liability for groups using international tort law has been made but the potential of introducing tort liability within international law has remained unexplored. The reason for this neglect is unclear but what is apparent is that there seems to be a robust need to hold entities accountable for their conduct which falls short of criminal liability. While international tort law could be a solution for

solving the accountability gap for LAWS, the dearth of scholarship makes this unlikely. However, tort law is perhaps the best solution for solving the accountability gap because it offers the possibility of regulating activities that are inherently dangerous but valuable and can offer compensation for wrongs that have been committed. While there is an established method for dealing with war criminals and holding them individually accountable, there is no such process to have states held accountable for their war torts. LAWS could be the best test case to see if war torts could be a valuable complement to international criminal law (Crootof 2016: 1352-1353).

3.7. HUMAN DECISION-MAKING AND MEANINGFUL HUMAN CONTROL

Regardless of the current and future capabilities of LAWS, the legal requirements of IHL seem to necessitate the presence of a reasonable human mind when making key decisions relating to IHL obligations. The need for human judgement as a legal requirement must be separated from the issues of morality. When states and the international community have considered life and death decisions being delegated to non-humans in the past, it has always been difficult to reconcile this concept with either moral or legal principles. However, these two concepts are often distorted or combined by commentators and courts (Beard 2014: 665).

While some commentators insist that there will always be a human in the loop, there is no guarantee that it will equate to the same standard of human responsibility, supervision, and control as if a human alone were performing the task. Furthermore, as weapons become increasingly more automated, it would become increasingly unfair and unrealistic to presume that a human will remain accountable for its actions. As seen in weapons equipped with low levels of automation, humans remain vulnerable to automation bias. Automation bias is best explained using two examples of current automated systems deployed by the US armed forces (Grut 2013: 14).

The first concerns an incident that took place in 1988 which involved a US Navy ship, the USS Vincennes, shooting down Iran Air flight 655. The ship was using the Aegis weapons system which was designed to defend the ship against attacks from the air. This system can be set to different modes of operation with differing degrees of

automation and the ability of a human to intervene at any stage and override the computer. During the course of events which transpired on July 3, 1988, the radar system picked up the airplane. All the indicators, including the speed, course, radio signal, and radar broadcast from the plane indicated that it was a passenger airplane. However, while operating in semi-autonomous mode, the Aegis misidentified the jet liner as a F-14 fighter jet belonging to the Iranian Air Force. While the evidence indicated the contrary, not one of the crew members were prepared to challenge the conclusions of the Aegis system; they trusted the system and authorised the engagement. Only after the airliner had been shot down did the crew realise that they had targeted a civilian aircraft, killing 290 people. The Aegis system was designed to operate in open skies in the North Atlantic, not the crowded skies in the Gulf. This is one of the clearest examples of automation bias due to the sheer amount of information outside of the Aegis system indicating that it was a commercial airliner.

The other incident of automation bias is not as clear cut. It occurred with the invasion of Iraq in 2003 when two allied jets were misidentified as incoming Iraqi missiles and subsequently shot down by US Patriot missiles (Grut 2013: 14).

The argument of meaningful human control makes sense on the surface; however, given the examples of the failure of human oversight, the above examples highlight that this might not be a realistic possibility. Add to this the possibility of battles being fought at machine speed and this would automatically eliminate the possibility of human oversight.

3.8. CHALLENGE: IHL IS NOT SUFFICIENT TO DEAL WITH LAWS

According to Wagner (2014: 1386), a number of authors have contended that IHL in its present form is inadequate to deal with the issues that LAWS present. Moral discourse and law have notoriously lagged behind technological development, especially in the realm of warfare. This, however, does not mean that morality and law are irrelevant and should be disregarded. Neither does it mean they are likely to be an impediment to technological advancement (Lucas 2013: 1277). In this context, the development of IHL has generally failed to keep up with technological advances in

warfare. While this is not a unique characteristic of IHL, as it could be argued that it is applicable wherever a legal regime interacts with advancing technology, the issue remains that, if LAWS are to be used in combat, they must be able to adhere to the principles of IHL (Foy 2014: 56). However, history has shown that IHL has demonstrated the capacity to adjust its rules in order to rise to the challenge posed by new weapons systems. IHL is comprised of broad rules and principles that apply to a wide variety of weapons systems as opposed to focusing on specific technologies and is therefore capable to respond to the challenge LAWS present, despite the multitude of differing opinions with regards to the interpretation of the existing rules and their capacity to deal with LAWS (Wagner 2014: 1386).

3.8.1. Defining Concepts

It is important to differentiate between the concepts of weapons, means of warfare, and munitions. LAWS, for example, are a means of warfare which comprises all the components of the weapons and munitions, including hardware and software. The defining feature of LAWS in this context is the use of AI software (Boogaard 2015: 247-250). It is important to clarify that neither the term weapon nor the terms means or methods of warfare are defined in IHL or the relevant legal instruments. Without this legal definition, we are dependent on a collection of stable and recognisable classifications that mould these terms. The linguistic roots of the word weapon indicate that it is an object intended or used for imposing harm or causing damage as well a tool for defending oneself or acquiring an advantage in a fight or a challenge. As such, it would be illogical to contemplate the features of a weapon without understanding the framework of how it will be used (Liu 2012 634-635). The ICRC defines weapons, means, or methods of warfare as:

> *The words "methods and means" include weapons in the widest sense, as well as the way in which they are used. The use that is made of a weapon can be unlawful in itself, or it can be unlawful only under certain conditions. For example, poison is unlawful in itself, as would be any weapon which would, by its very nature, be so imprecise that it would inevitably cause indiscriminate damage. It would automatically fall under the prohibition of Article 57. However,*

a weapon that can be used with precision can also be abusively used against the civilian population. In this case, it is not the weapon, which is prohibited, but the method or the way in which it is used (ICRC 1987a).

3.8.2. Targeting Law

With regards to the applicable legal framework, there is no need to treat LAWS differently compared to any other means of warfare. However, the argument could be made that, because LAWS use AI to decide on what targets to engage and how to engage them, they should be treated differently compared to human involvement in making these decisions in a traditional warfare setting. The challenge this poses is that it will ultimately be humans who decide to deploy these AWS, with the consequence being that the rules for targeting in IHL would apply to LAWS as they do to other current means of warfare (Boogaard 2015: 255), meaning LAWS would have to adhere to the principles of distinction and proportionality.

3.8.3. The Adaption of Law to Technical Advancement

Technology and law have coexisted in tension where over time a new equilibrium has been established. An example of this is the attempt to ban the crossbow from medieval battlefields because it was seen as being hateful towards God and because ordinary soldiers could use it to kill knights. In contrast to complete bans on classes of weapons, technological changes and shifting expectations and experiences have brought about necessary legal reforms. A modern example is how the targeting of whole cities during World War II necessitated changes to targeting law which required the identification of specific targets within an area that contained civilians or civilian objects such as a city, a town, a village, or a similar constellation (Newton 2015: 10).

The campaign for a moratorium on all LAWS must be assessed against the backdrop of politics and personalities. Some of the people involved in the campaign are the same people who were involved in the 1997 Ottawa Convention which banned the production, transfer, and use of APL. The spread of improvised explosive devices (IEDs), which are the functional equivalents of landmines, and the impact they have

had on both combatants and civilians, provides insight and a warning into the unintended consequences a total ban on LAWS could have. Total bans have seldom been effective in changing the way states behave simply because they can have a disproportionate advantage in disregarding the rules (Newton 2015: 9). International law has demonstrated the ability adapt to changes in weapons, munitions, and the means and methods of warfare and while LAWS do pose a number of unique challenges, these challenges can be overcome. There is also the warning in the example of IEDs: even if there are legal instruments in place prohibiting weapons, munitions, and the means and methods of warfare, certain groups can disregard this and act in their own interests.

3.9 CONCLUSION

This chapter has categorised and explained the major challenges with regards to the legal aspects of LAWS. It is conceivable that LAWS would not be declared outright illegal. However, their use may be far from legal under the current legal framework. LAWS would need to adhere to the principles of proportionality and distinction which would be very difficult to implement. LAWS also pose the problem of a unique accountability gap that would need to be closed within current legal enforcement mechanisms to ensure a perverse incentive to develop LAWS to skirt individual accountability is not created. Meaningful human control has been suggested as a way of closing this gap, but little evidence exists that this would be a foolproof way of solving these issues. While it has been argued that international law in its current form would not be capable of dealing with the challenges LAWS pose, it has been shown that international law is capable of adapting to the changing environment, even if it is in retrospect.

4. CHAPTER FOUR: LAWS AND THE POTENTIAL FOR ARMS CONTROL

4. 1 INTRODUCTION

Attempts at regulating LAWS have received international attention in discussions through the framework of the CCW. While no common understanding yet exists of what LAWS actually are, the current debates have centred on whether a moratorium or ban should be placed on the development and fielding of LAWS in combat. While the discussions are likely to continue for some time to come, a tangible next step would be to conduct a legal review of weapons systems. In addition, the US has suggested an intermediate step which would entail the compilation of a document that would comprise of the best practices in conducting a comprehensive weapon review on LAWS (Meier 2016: 120).

4.2 ARMS CONTROL

Arms control is defined as any agreement either directly between two parties or multilaterally within the in the broader international community that is aimed at directly limiting the weapons that could be used in warfare (Hanson 2012: 173). The classical definition provided by Bull (1961: vii) states that "[a]rms control is restraint internationally exercised upon armaments policy, whether in respect of the level of armaments, their character, deployment or use". The forms that arms control can take vary between a formal process, including treaties or other forms of binding agreements, to an informal practice between parties. The above processes can be conducted either as multilateral, bilateral, or even unilateral processes, however, the crucial ingredients are the willingness of the parties to collaborate with each other to attain mutual security interests which might either be self-serving for the cooperating states or shared by the international community. In practice, arms control is applied to both WMDs as well as conventional weapons with the focus more often on WMDs as these weapons have the greatest tendency to inflict large scale indiscriminate destruction and suffering on large volumes of peoples. (Hanson 2012: 173). Therefore,

arms control aims to initiate a formal process between parties that results in a limitation being imposed on the acquisition, development, or use of a certain class of weapon.

4.2.1 Incentives to Adopt Arms Control

The first incentive to adopt arms control is that if the number or kinds of weapons that states can possess is limited, this results in the direct reduction in the possibility of war breaking out between them. As a result, arms control is directly responsible for the reduction of, if not complete removal of, the negative impact of the security dilemma. A security dilemma exists when a state or groups of states are uncertain about the intentions of their adversaries. This perceived uncertainty can drive a state to arm itself in order to prepare for the possibility of defending itself in the event of an attack. However, this often has the opposite effect in that the state or groups of states that feel uncertain and interpret the build-up of capabilities in another state or group of states as a possible indication of attack may drive these states to further increase their military preparedness. This reinforces suspicion and leads to further escalation, which can result in an inevitable spiral and escalation of both arms acquisition and suspicion of the other state or states (Hansen 2012: 174). The security dilemma is used to explain peacetime spirals of increasing military preparedness and escalating political tensions that generate arms race type behaviour as well as the resulting inabilities in the system (Wong *et al.* 2020: 15).

Wars can occur as a consequence of the fear over the military power of another state or states and, as such, any agreements that open up communication and limit the number of types of arms can foster greater transparency of intentions of the other state or states. This becomes a confidence building measure. As wars can also result in multiple casualties, there is an increased focus on the humanitarian motivation to reduce such casualties. Therefore, humanitarian aims are often considered as the main driving force in the disarmament efforts of modern arms control (Hansen 2012: 174). Thus, arms control can be used to reduce both the security dilemma and the humanitarian impact of conflict.

4.3 THE ROLE OF ARMS CONTROL FOR LAWS

Arms control processes cannot exist in a vacuum as they require a supportive environment which enables them to be effectively implemented, especially if the agreements are merely voluntary and not legally binding. This is even more important when there is increased competition in and for economic, political, and military advantages; weakened multilateralism; and tense relations between great powers. These factors all influence the ability of certain measures and tools to affect the outcome. In relation to controlling weapons, arms control can offer four objectives. The first is stability which aims to remove any incentives for first strikes as well as preventing accidental war and reducing the risk of military escalation through greater transparency and predictability. The second objective is that of safety which aims to reduce the risks associated with military operations. The third is legality which aims is to ensure compliance with international legal obligations, including IHL and IHRL. The fourth is efficacy which aims to provide enough incentives and, with that, a better chance of successful controls to be deployed and the desired behaviour of states to be produced. The most applicable of these to LAWS would be stability and safety (Persi Paoli *et al*. 2020: 2). These two objectives of arms control are particularly pertinent given that LAWS could generate a classical security dilemma, and this is seen as even more likely given the potential for an arms race between China and the US in autonomous systems (Wong *et al*. 2020: 61). Given the tensions between the US, Russia, and China, who have all committed to developing leading AI capabilities, there is a risk that an arms race will develop. The benefits of engaging in some form of arms control for LAWS would reduce these risks.

4.4. CURRENT REGULATORY FRAMEWORK

Specific international agreements exist to regulate weapons that have been classified as inherently problematic weapons. These include poisonous gases, expanding bullets, landmines, and blinding lasers (Grut 2013: 9). As there are no current treaties that ban the category of AWS in their entirety, sates have the obligation to prove if a new weapon should be banned by the current rules of IHL. This responsibility is rooted in the limitations imposed on parties with regards to the choice of their means of

warfare. These restrictions exist in numerous treaty provisions and customary IHL. The initial overall responsibility that states have when considering if a new weapon is legal is to assess if, by its very nature, the new weapon could cause unnecessary harm or needless suffering. Though a precise definition in this context is elusive, practically it is the trade-off between the tactical advantages of the weapons system versus the harm the weapons system can inflict on combatants. For example, bullets that expand or explode on impact, wounding the combatant so severely that they have slim to no chances of survival, or lasers that cause permanent blindness, are both illegal based on the above criterion. In these cases, it is not the weapons system itself that has been outlawed but rather the impact of the type of munition used on enemy combatants which does not necessarily take into account the impact these munitions would have on the civilian population if they were not participating in the hostilities directly. An example of this is APLs. While these landmines had a dramatic impact on civilian populations, it was in fact the impact it had on soldiers that motivated states to decide to ban their production, sale, transfer, stockpiling, and use.

Using these examples, it is clear that automating the target selection and attack would have no impact on addressing the question of causing superfluous harm or unnecessary suffering. LAWS would most likely be equipped with similar weapons and munitions that are in use today and operate in a similar manner. Therefore, it can reasonably be expected that LAWS will comply with the provision that prohibits means and methods of warfare which inflict unnecessary harm or needless suffering (Boogaard 2015: 257-258).

The second critical obligation of states with regards to the legality of new weapons would be to ensure that the weapons observe the principle of distinction. In fact, it could be argued that there is no more important principle underlying IHL. The principle of distinction is simply the obligation to distinguish between enemy forces and their installations and civilians and their objects as well as other groups of protected persons, including peacekeepers, medical personal, and sick and wounded soldiers. This principle also precludes indiscriminate attacks and weapons that cannot be directed towards a specific target. Examples of these types of weapons include the V2 rocket used by Nazi Germany and the Scud missile used by Saddam Hussein's Iraq.

Both lacked precision targeting and the ability to limit the damage caused to people and property.

However, even if a weapon is classified as not being intrinsically indiscriminate and if it does not by nature cause unnecessary harm or needless suffering, other restrictions still apply to the weapon under consideration. These include the assessment of whether or not the weapon can be used indiscriminately, unlawfully putting civilian populations at risk, and precluding a direct attack on civilians with what would under normal circumstances be a legal weapons system. Therefore, to assess if LAWS are inherently indiscriminate, the need to assess the weapons system's ability to correctly classify civilians is of paramount importance. The ability to perform this classification is necessary in order to comply with both the distinction principle as well the principle of proportionality (Boogaard 2015: 258).

If during a weapon review the weapon is deemed to be legal, the use of this weapon will still need to comply with additional restrictions. This introduces the obligation to minimise civilian casualties. Proportionality is also part of the obligation to take necessary precautions when attacking, which includes the preparation for attack as well as attacks that are ongoing, as this responsibility is directed to those that plot or elect to initiate an attack. As the overall rules of weapons law and the precise rules of targeting contained in IHL would apply to LAWS, it can be reasonably presumed that the obligation to take the necessary precaution is directed not at the weapons system itself but rather at the human controllers. However, LAWS would still have to comply with other specific protective obligations which may prove challenging.

An example of this would be the need to differentiate between a normal active soldier, a soldier that is *hors de combat*, and medical personnel carrying a weapon for self-defence. This extends to vehicles and the need to differentiate between a standard armoured personnel carrier and one that is used as an ambulance. The weapons system would also need to differentiate between civilians who participate in hostilities for only a limited time and the civilians who did not participate at all – which may prove to be a challenging task. The complexity furthermore extends to the duty to not attack someone who has undoubtedly conveyed their intention to surrender (Boogaard 2015: 259-260). Therefore, it is critical to assess both on a state level as well as amongst

states if new weapons are legal and to what extent they comply with current international law.

4.5. THE CONVENTION ON CERTAIN CONVENTIONAL WEAPONS

A few mechanisms can be used when assessing new weapons and the CCW is one of these mechanisms. The CCW exists to prevent or curtail the use of specific conventional weapons which are believed to be either extremely injurious or weapons which have an indiscriminate impact (Sauer 2016: 10). Currently 125 states are party to the agreement with an additional four signatories (UNODA 2021a). The CCW consists of a number of protocols that regulate specific types of weapons which form the framework of the convention. For example, Protocol IV banned BLW pre-emptively. When weapons are deliberated on in the CCW the outcome tends to result in a treaty. However, the process failed to produce a result on cluster munitions which were deliberated on in 2011. This resulted in the Convention on Cluster Munitions, which was created outside both the CCW and UN, and is the sole international instrument to specifically regulate this type of weapon.

Autonomous weapons have climbed the top of the CCW agenda in recent years (Sauer 2016: 10) and the first informal expert meeting of the CCW on LAWS took place between 13 and 16 May, 2014 with the mandate to discuss the emerging technologies surrounding LAWS. The technical, legal, military, operational, and ethical aspects related to LAWS were discussed. During the session on the legal aspects of LAWS, discussions were had around whether LAWS could comply with internal law as it exists, specifically IHL, including customary international law, the Martens Clause, and the 1949 Geneva conventions. Both legal experts and states emphasised the need for LAWS to comply with IHL. However, there were differing opinions on whether that would be at all possible, especially given the state of the technology. The requirement to ensure that weapons systems are indeed able to be used in accordance with international law through conducting a legal review on LAWS was also discussed. It was decided that the application of a weapons review, including Article 36 of AP I, would be valuable (Meier 2016: 120-121).

The second informal meeting of experts on LAWS convened in April 2015 where the discussions continued exploring the same topics. In a session dedicated to the challenges faced in terms of LAWS compliance with IHL, the legal weapons review process was deliberated. It was decided in a formal meeting of experts that the compliance of LAWS with IHL would be dependent on the nature of the weapons system, the specific mission, as well as the setting in which the weapons system is used. The added intricacy introduced by increasing the levels of autonomy of systems meant that the predictability of these systems would be constrained and that deploying a system with unpredictable outcomes would be in contravention of IHL. Therefore, a comprehensive weapons review is needed to ensure that the deployment of LAWS is compliant with IHL (Meier 2016: 122).

During the Fifth Review Conference of States Parties to the Convention on Certain Conventional Weapons in 2016 it was agreed by the High Contracting Parties that they should create a Group of Governmental Experts (GGE) on LAWS with the obligation to deliberate on the emergent technologies in the domain of LAWS (UNODA 2021b). This culminated in the adoption of 11 guiding principles by the High Contracting Parties to the CCW in 2019 based on the recommendation of the GGE on LAWS (UNODA 2021b).

4.5.1. The CCW Guiding Principles on LAWS

> *Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*
>
> *It was affirmed that international law, in particular the United Nations Charter and International Humanitarian Law (IHL) as well as relevant ethical perspectives, should guide the continued work of the Group. Noting the potential challenges posed by emerging technologies in the area of lethal autonomous weapons systems to IHL, the following were affirmed, without prejudice to the result of future discussions:*

(a) International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems;

(b) Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system;

(c) Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole;

(d) Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control;

(e) In accordance with States' obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law;

(f) When developing or acquiring new weapons systems based on emerging technologies in the area of lethal autonomous weapons systems, physical security, appropriate non-physical safeguards (including cyber-security against hacking or data spoofing), the risk of acquisition by terrorist groups and the risk of proliferation should be considered;

(g) Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems;

(h) Consideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations;

(i) In crafting potential policy measures, emerging technologies in the area of lethal autonomous weapons systems should not be anthropomorphized;
(j) Discussions and any potential policy measures taken within the context of the CCW should not hamper progress in or access to peaceful uses of intelligent autonomous technologies;
(k) The CCW offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems within the context of the objectives and purposes of the Convention, which seeks to strike a balance between military necessity and humanitarian considerations (CCW 2019).

However, some states, including Russia, have no appetite for generating new international law in CCW, maintaining that there are already established mechanisms within international law that can be used to address the concerns regarding the development and use of LAWS. Other states like Germany consider the 2019 report and its guiding principles as a politically binding regulation. On the other hand, the Campaign to Stop Killer Robots has been vocal in criticising the CCW as a result of the glacial pace of progress and the insignificant outcomes and low level of ambition in attempting to regulate LAWS despite public opposition. The Campaign to Stop Killer Robots is calling for some form of immediate negotiation leading to a biding legal instrument rather than continuing the vague and redundant focus on frameworks and principles (Sauer 2020: 236-237). A further two sessions of the CCW meeting on LAWS are scheduled for 2021 and the CCW sixth review conference is scheduled for December 2021 (Campaign to Stop Killer Robots 2021). While the guiding principles are a start in dealing with the potential regulation of LAWS, there should be some form of renewed urgency given that LAWS have now been used in combat. However, it will remain to be seen if the CCW responds accordingly.

4.6. THE REVIEW OF NEW WEAPONS

Limitations are placed on the choice of the means and methods of warfare for combatants. As one of the instruments available, the ICRC published a guideline on the legal review of new weapons, means, and methods of warfare in 2006. The guide

was issued to provide assistance to states in order to improve or establish a processes to investigate whether new weapons would be legal in accordance with their obligations that stem from Article 36 of AP I to the 1949 Geneva Conventions as described by the ICRC (ICRC 2006: 1):

> *This Guide aims to assist States in establishing or improving procedures to determine the legality of new weapons, means and methods of warfare in accordance with Article 36 of Protocol I Additional to the 1949 Geneva Conventions. It was prepared further to an expert meeting hosted by the ICRC in January 2001 and the Agenda for Humanitarian Action adopted by the States Parties to the Geneva Conventions at the 28th International Conference of the Red Cross and Red Crescent. The Agenda for Humanitarian Action commits States to ensure the legality of all new weapons, means and methods of warfare by subjecting them to rigorous and multidisciplinary review* (ICRC 2006: 1).

Each State Party to Article 36 of AP I has to fulfil the requirement to assess whether new weapons, means, and methods of warfare that are developed, studied, adopted, or acquired would, in part or in the whole, be banned by international law. All states should take an interest in evaluating the lawfulness of new weapons, irrespective of whether they are a party to AP I. Through the assessment of the legality of new weapons, sates have the opportunity to ensure that they are proficient in the conduct of hostilities in compliance with their international responsibilities, especially in light of rapid technological developments (ICRC 2006 :1).

The ICRC (2006: 1) guideline highlights that the legal review is applicable to weapons in their broadest definition, including the way that the weapons would be used, as weapons cannot be evaluated separate from their anticipated use. The structure for the legal review is the international law that applies to the state, especially IHL. This comprises of the treaty itself as well as the customary restrictions and prohibitions on particular weapons as well the overall IHL rules applicable to the use of all weapons as well as the means and methods of warfare. It goes on to emphasise that general IHL rules include:

> *the rules aimed at protecting civilians from the indiscriminate effects of weapons and combatants from unnecessary suffering. The assessment of a weapon in light of the relevant rules will require an examination of all relevant empirical information pertaining to the weapon, such as its technical description and actual performance, and its effects on health and the environment. This is the rationale for the involvement of experts of various disciplines in the review process* (ICRC 2006: 1).

However, the limitations of this process are obvious. Firstly, no international standards exist for conducting a weapons review other than the ICRC guide and legal commentary. Implementation of an Article 36 review can also suffer from states focusing only on the legal risk, rather than the potential ethical or humanitarian concerns that these new weapons pose. Furthermore, LAWS are of international concern and therefore decisions about adopting LAWS should not rest only with individual states. National reviews could be badly designed or too narrow in scope given the complexity of the challenges LAWS pose and it would be ill conceived to delegate these decisions to individual states rather than within multilateral discussions. The CCW exists as a response to the limited effectiveness of weapons reviews (Article 36 2016: 2). Finally, only a limited number of states actively engage with the weapons review process and currently no prohibitions on the development, adoption, or acquisition of new weapons systems using this mechanism have been set out (Grut 2013: 10).

Expanding this review to additional states would require a significant development of capacity which is unlikely to be achieved (Article 36 2016: 2). However, it is problematic to contend that Article 36 actually forms part of customary international law, even though the ICRC has insisted that the evaluation of the legality of new weapons is a responsibility for all states. Nevertheless, with the dearth of detail present in Article 36 and the absence of actual state participation, it is not very clear what the compliance with the Article 36 review will look like given the circumstances described (Grut 2013: 10). While the Article 36 review does make it possible to identify the challenges new weapons pose, LAWS present a more comprehensive challenge to the existing norms and legal provisions. Therefore, when discussing potential arms

control agreements for LAWS, multilateral and legally binding processes like the CCW are the preferred avenue.

4.7 NORM SETTING

It is also useful to consider the building of norms as a potential method of influence when attempting to regulate weapons. Successful norm building requires active and committed norm entrepreneurs in the form of advocacy groups, the individuals or groups affected, etc. who initially raise awareness of the issue and then construct a framing that resonates which enables the mobilisation of the audience and applies pressure to norm addressees and, finally, the choice or creation of a favourable institutional setting. The resulting strategies can then be broken down into their individual actor related, issue related and institutionally related components (Rosert & Sauer 2021: 7). The following discussion will highlight examples of where these techniques have been used successfully in the past to regulate APL and BLW.

4.7.1. Case Comparison

Scholars of arms control in IR have shown that humanitarian advocacy campaigns have wielded influence and have demonstrated the ability to both shift sentiment and arrive at agreements, even when these are not in the interests of advanced military powers. The scholarship has even highlighted which strategies these organisations can adopt to make their desired outcome more likely to succeed. As it is generally agreed that the Campaign to Stop Killer Robots as a representative for civil society is responsible for bringing weapon autonomy to the attention of the UN's arms control processes, this provides the opportunity to examine the campaign's strategy against two other previously successful campaigns conducted against APL and BLW. These are useful comparisons as both are often referred to in the discourse on LAWS because the possibility exists that LAWS would be incompatible with IHL in the same way that both BLW and APL are and both of these weapons were successfully regulated (Rosert & Sauer 2021: 6). It is important to examine the process as well as the outcomes of the attempt to regulate both of these problematic weapons as it can

give valuable insight into the process as well the potential outcome which can be extrapolated for LAWS.

4.7.2. Ban on Anti-Personnel Landmines

The ban on APL was framed as the first main achievement in disarmament based on humanitarian grounds. Initially the CCW did not pass a complete ban in the Amended Protocol II but rather distinguished between the classes of anti-tank and APL, including differentiating between smart and dumb mines. This outcome was not ideal and resulted in both states and NGOs seeking a stronger outcome outside of the UN framework. They established what become known as the Ottawa Process. In 1997 the Convention on the Prohibition of the Use, Stockpiling, Production, and Transfer of Anti-Personnel Mines and on their Destruction was passed by the participants. It is considered to be one of the most successful treaties as the social influence attached to this process has induced even non-parties to the convention to comply. The Ottawa Treaty, as the convention is informally known, has virtually stopped the manufacture, trade, and use of APL; it is the driving force behind the removal of millions of APL throughout the globe; and it has provided assistance for its many casualties (Rosert & Sauer 2021: 12).

The framing of the ban on APL had to, from the outset, redefine three characteristics of the common perceptions of APL. The initial challenge was that, even though APL were thought to be cowardly and barbarous when they were introduced to the 19th century battlefield they had come to be regarded as a useful weapon of no real ill repute. The second challenge was that, through their defensive capabilities, APL were seen as legitimate weapons. Finally, because of their military utility, APL were thought of as indispensable, effective, and efficient weapons with undisputed universal use. Therefore, in order to force the issue, the depoliticised and neutral discussions on APL necessitated the transformation of the discourse into a humanitarian one. This was achieved through a broad appeal to human emotions by placing the victims at the centre, as well as triggering the human security frame, highlighting the socio-economic costs, and drawing comparisons to WMD (Rosert & Sauer 2021: 12). In order to

achieve the desired outcome, the norm entrepreneurs had to change the way APL were framed, thought of, and discussed.

Civilians were framed as being the most likely victims of landmines despite the data proving otherwise, thus demonstrating that these weapons are inhumane and violate the IHL principle of distinction. The extent of the issue of civilian casualties of APL was unknown at the time but the data collected showed not only the extent of the issue but also the geographic distribution of their use. Photographs and stories of victims offset the neutral political discourse and anonymous data points. The campaign also highlighted the socioeconomic cost of APL through detailing the reduced agricultural activities in contaminated areas, the cost of clearing APL, the financial burden on healthcare systems and families caring for victims, as well as victims' reduced capacity to earn an income. In addition, the campaign highlighted that studies had shown that the military usefulness of APL was limited given their limited utility as an offensive weapon as well as the limited impact on the outcome of conflicts. The issue of APL also happened to neatly align with the shift from state-centric security to that of the security of the individual that occurred at the end of the Cold War. Thus, APL were not seen as defenders of the borders of states but rather as gruesome threats to individuals. Finally, there was the claim that APL and WMD share three similar features: the ability to kill indiscriminately, the threat to masses of people, and that control over them can easily be lost. This ultimately resulted in a corresponding normative demand that APL, like WMD, were in fact not usable and should be completely prohibited (Rosert & Sauer 2021: 12-13). By seizing the humanitarian agenda, the norm entrepreneurs were able to show the multi-dimensional human cost to civilian lives of APL and put this cost foremost in the minds of ordinary people.

An extensive number of norm entrepreneurs, which included HRW, the ICRC, and a few governments, dedicated their focus on banning APL. The ICRC had added APL to a list of weapons that could be considered inhumane after the end of World War II that they desired to control. This desire ultimately developed into the weak Protocol II of the CCW in 1980. After a number of years a few minor but seasoned NGOs, namely Physicians for Human Rights, Asia Watch, as well as the Women's Commission for Refugee Women and Children really kick started the movement to prohibit APL by utilising public petitions and reports. In 1992 the International Campaign to Ban

Landmines (ICBL) was co-launched by HRW. It grew rapidly to over a thousand members comprised of both international and national NGOs, including the UN Department of Humanitarian Affairs and UNICEF. The list of governments that supported the movement included the US, Belgium, Canada, and France (Rosert & Sauer 2021: 13). It was this collaboration between different organisations and states working to achieve the same goal that was successful in mobilising the agenda in the public mind which resulted in the banning of APL, even if it was outside of the purview of the CCW.

4.7.3. Ban on Blinding Laser Weapons

The ban on BLW was implemented in 1995 at the First Review Conference of the CCW as Protocol IV. This was considered to be a remarkable achievement for a number of reasons. At the time the CCW negotiations only resulted in restrictions on weapons and this was the first outright ban of a weapon. This was also the first time that both the use and transfer of a weapon were banned. In addition, this was the first weapon to be banned pre-emptively as in the past restrictions on weapons had been implemented retrospectively, proving that it is possible to pre-emptively ban new weapons. It was the first modern weapon to be banned on the principle of unnecessary suffering and with the protection of combatants in mind. Finally, BLW were banned despite a less intensive transnational campaign than was witnessed for APL and is currently underway for LAWS (Rosert & Sauer 2021: 9).

The interest in the military application of Light Amplification by the Simulated Emission of Radiation (LASER) dates back to the 1960s and approximately thirty countries' armed forces have explored the technology since then. The interest in LASERs stem from the fact that they can travel silently over great distances at the speed of light in a straight line. Thus, LASERs are faster; more accurate than any projectile; require no logistical effort; and only need energy, not ammunition. Their current military applications include target designation for guided munitions, range finding, training simulation, and communications. LASERs also possess the capacity to destroy optical sensors and, in some cases, even non-optical material. Therefore, LASERs possess the capacity to soft kill a target by incapacitating armour through destroying its

sensors. The side effect of this is that they can damage the eyes of the human crew as either collateral damage or as part of the desired effect. However, despite this humanitarian concern, LASERs were deemed acceptable and even desirable.

BLW share a number of parallels with LAWS in that it is a technology with dual military and civilian use, it is a new weapon, a number of militaries have explored its potential use, and it has a number of possible uses of autonomy in the military. However, their potential impact on the on combatants and civilians is not as clear cut.

LASERs were included in the Ad Hoc Committee of the ICRC Diplomatic Conference's international agenda in the in the 1970s and were considered to be futuristic weapons at the time. This futuristic slant made it less of an urgent matter given other inhumane weapons of the time such as napalm that was already causing terrible suffering. NGOs shifted the tactics to address the issue of BLW on their own without being categorised. While the weapon might have been new, its effect of blinding humans was not (Rosert & Sauer 2021: 10).

The strategy that was followed in an attempt to ban BLW was to use the idea of blinding humans while in tandem dispelling the opinion that being blinded is preferable to being killed. The advocates for banning BLW initiated three key frames to advance their cause. The first was that of the non-compliance with IHL given they are weapons that cause unnecessary suffering. Not only could soldiers permanently lose their sight with no possibility of reconstruction, it could be extremely painful and, in some cases, the eyeballs could explode. It would also lead to a permanent disability that would cause psychological trauma, depression, and social isolation. The second frame was the fear of blindness. Sight provides the majority of our sensory stimulation and it is possible for humans to imagine being blind. Societies also have an understanding of this disability, as had been seen during World War I as a result of mustard gas and the Vietnam War with cluster munitions. The ICRC expected there would be a rise in cases of combat stress disorder as well as a public outcry on the impact of BLW. Finally, the impact that combat veterans would have on socioeconomic spheres was considered. It was argued that veterans who have been blinded would be limited in their ability to participate economically and would overload the medical and social infrastructure (Rosert & Sauer 2021: 10-11). Here, as with APL, the focus on the

humanitarian cost to the individuals and society would be too large and this should be reason enough to ban the weapons.

Initially the attempt at a ban was state-led with Switzerland and Sweden wanting to implement a ban. The US and the Soviet Union followed suit with their 1989 bilateral agreement on withdrawing from dangerous military activities, which included BLW. The ICRC then proceeded to adopt the issue in 1989 and rather than mobilising the public opinion or actively embarrassing states, they assumed the role of advisor, providing expertise in order to persuade governments that BLW are inhumane. This resulted in four expert meetings on the issues BLW would generate between 1989 and 1991. As more parties came on board, a coalition of supporters for the ban emerged with NGOs like HRW, which framed BLW as a human rights concern. France and Sweden responded to the ICRC's appeal and between 1995 and 1996 assembled a Review Conference of the CCW and in its agenda included BLW. It was here that the ICRC's work on educating governments on the impact of BLW paid off. The US changed their position from initially opposing the ban to actively supporting it by leading the way with a national non-use policy. Nevertheless, there were still disagreements, especially on whether to ban just the weapons or the methods of warfare. France and Germany, for example, were looking for a ban on the method of warfare. However, the US and UK wanted a limited ban that included only weapons that were intended to inflict permanent blindness, not accidental injuries. After a closed meeting amongst the participants, the US position succeeded and in October of 1995 Protocol IV to the CCW was adopted (Rosert & Sauer 2021: 11-12). After a number of years on the agenda, BLW were banned before they were used. However, with the use of LAWS now documented, it is already too late to pre-emptively issue any form of proactive arms control for LAWS.

4.8. CONCLUSION

Arms control plays a vital role in establishing and maintaining peace amongst states given the role it can play in actively reducing the security dilemma through enhancing transparency and reducing uncertainty. There is a very real possibility that AI will induce arms racing behaviour, especially between Russia, China, and the US in the

development of this technology and that is likely to be weaponised. Therefore, there are real incentives to maintain some form of global stability by implementing arms control when it comes to LAWS. There are a number of avenues available for the potential regulation of LAWS through arms control agreements as well as exiting responsibilities that states have with regards to reviewing new weapons. However, LAWS have been under review in the CCW for a number of years already, starting with informal meetings between 2014 and 2016 and progressing within the framework of the GGE since 2017. This has resulted in a set of guidelines, but this is far from a treaty directly addressing the topic. In addition, the national new weapon review obligation as part of Article 36 of AP I is not sufficient to tackle such an issue on a global scale. With the first use of LAWS in battle now documented, the genie is out of the bottle. However, the successful regulation of both APL and BLW demonstrates that it is indeed possible to regulate undesirable weapons, even if it is outside of the CCW process.

5. CHAPTER FIVE: THE ETHICS OF LAWS

5.1 INTRODUCTION

The just war tradition has evolved over centuries and is thought of as a living tradition as the appreciation of what is just has significantly evolved, opening up the tradition to change. In western traditions, just war conversations started in ancient Israel where rules limited the use of violence and even demonstrated more compassion to fruit trees than to people. The ancient Romans and the Greeks showed little compassion for barbarian civilians. In the Middle Ages, Christian scholars had written about protecting non-combatants, but with some pragmatic concessions. The current debates reflect not only the nature of the technological ability to be more precise in the targeting, and thus reducing, the impact on others not involved in the conflict but also our changing attitude towards human life (Morkevicius 2014: 7-8). LAWS occupy a shade of grey in both the legal and ethical dimensions of war. The purpose of this chapter is to highlight the key ethical considerations and challenges LAWS bring to the ethical dimension of conflict.

One of the biggest legal challenges discussed in the previous chapters is that of the ability for LAWS to make valid targeting choices. However, making a decision around the validity of a target requires a judgement call of what is permissible. Determining what is permissible, regardless of what framework it is conducted under, is fundamentally an ethical and moral decision (Umbrello *et al*. 2020: 277-278). Scharre (2018b) recounts a personal experience where he, as part of a US sniper team that was deployed to Afghanistan, had moved into position under the cover of darkness only to find at daybreak that the position was not as well disguised as had been hoped for. The village became aware of their position and a little girl, estimated to be between five and six, was sent to scout their position under the pretence of herding some goats. The girl was circling their position reporting back on a radio and not long after she left Taliban fighters emerged and engaged the sniper team. After a brief battle with the Taliban the sniper team withdrew from their position. At the debrief, the sniper team discussed what they would do next time they came across a similar situation. One of the things they did not discuss as an option was shooting the girl scouting their position. However, under the Laws of War it would have been legal to kill her; no age

is assigned to combatants because your combatant status is determined by your actions alone. This girl was actively scouting for the enemy, thereby participating in hostilities, which makes her a combatant whether she was doing so voluntarily or if she was forced into it. This example highlights the ethical issues that have to be confronted with the development of LAWS, including the ethics or morality that should be included as a starting point.

5.2 UNIVERSAL ETHICAL FRAMEWORK

One of the issues concerned with the development of ethics for robots is that philosophers do not agree on what theory of ethics should be used (Bogosian 2017: 592). Consequently, there is no agreed universal moral framework with competing viewpoints. Divine Command Theory (DCT) aims to coordinate moral action according to the divine, for example from holy scripture. Utilitarianism seeks to weigh the welfare generated by the various options and choose the one which results in the most favourable consequences for the largest number. Deontology aims to apply moral principles which are universal in nature to prescribe action. Lastly, virtue ethics looks to people to express the character of model moral personalities, for example Socrates, Christ, Epictetus, and Aurelius and direct their actions accordingly. However, these all have issues. DCT, for example, is very interpretive and there is no agreement on which words of which divine being are used. There is also no real epistemic basis for deciding between the alternatives. Utilitarianism suffers from the lack of a definition of what is good, which consequences are ethically applicable, and how to actually determine the quantity of good generated by an act. Deontology has the issue of generating unpalatable consequences, for example indicating to a Nazi soldier that one is concealing someone in the cellar because of duty to always tell the truth (Umbrello *et al*. 2020: 278).

In 2009 a survey was conducted that revealed that 26% of philosophy faculties preferred deontology, 24% preferred consequentialism, 18% preferred virtue ethics, and the remaining 32% preferred different approaches completely. If the belief existed that a particular approach to ethics is the best approach then the correct choice for an artificial agent such as a robot would be simple. However, the road to adoption is far

from clear for both institutions and society as a whole because researchers, governments, and companies are faced with difficult choices between the competing approaches and deciding which approach to ethics should be applied to artificial agents (Bogosian 2017: 592). This highlights the fact that there are many competing ethical viewpoints. The resulting debate as to the right ethical approach and the lack of a universally agreed framework puts the consideration of the ethical dimension of LAWS into very difficult terrain.

5.3 MORAL AGENCY

Moral agency is the idea that actors are capable of behaving in agreement with the principles of morality, that they have the ability to make moral choices autonomously, and can be held accountable for their actions. Therefore, moral agency inherently means that agents have autonomy, that they act independently from the will of others, and that their actions originate within them and reflect the ends which they want to achieve. This also means that it involves intention as actors intend to achieve the ends that result from their actions. Therefore, agents might be held morally responsible for the actions they commit to the degree that the outcome was indeed intended and was executed autonomously. This concept of moral agency is challenged by autonomous systems, especially in the case where the systems take care of the identification of a potential target and then rely on a human operator to execute the attack (Steele & Heinze 2014: 100-101).

As already discussed in this study, an example that highlights this dilemma is the case of the downed Iranian airliner where the crew of a US destroyer mistook it for an Iranian F-14 fighter jet. Despite evidence to the contrary, the crew trusted the decision of the system over their own judgement. In this case, who should be held accountable? As the misidentification by the Aegis combat system provides some partial excusability on the side of the human operators, what is interesting in this situation is that the misidentification was made by the system and that the mistake on the side of the humans was to trust the system instead of their own judgement. If this incident had happened without the use of the Aegis system, the human operators would have been responsible for not having taken reasonable precautions as they would have been

expected to have taken steps to guarantee that the target was undeniably a military threat. The result here is that with increasing reliance on these targeting systems with this type of autonomous technology the system itself becomes the reasonable precaution which then provides a form of moral defence between the actions of the human operators and the resulting consequences. This, in turn, allows the operators to believe that it was the machine that made the choice, thereby relieving themselves of moral responsibility when they are actually the only moral agents in the scenario (Steele & Heinze 2014: 104-105). Therefore, LAWS will not have moral agency but this lack of moral agency must not result in a conceptual moral buffer which isolates the human operator from their responsibility.

5.4 ETHICAL ROBOTS IN WARFARE

Robots are incapable of behaving morally or immorally or behaving more humanely than humans. However, humans are capable of behaving inhumanely (Sassóli 2014: 313-314). Arkin (2009: 31) has argued that LAWS would ultimately reduce collateral damage and the deaths of civilians because they will be able to better comply with the Laws of War than a human. Firstly, he argues, robots have the capability to be conservative in their approach as they have no need for self-preservation. With the drive for self-preservation removed, robots are freer to exercise caution with low certainty targets as the cost of misclassification is low and robots can even be utilised for self-sacrifice if need be. The deployment and usage of a wide variety of battlefield sensors would enable a clearer understanding of the battlefield in far more detail than would be capable with human sensory perception. Robots do not have any emotions that could impact their judgement and they will not be impacted by frustration, anger, or fear. Robots are immune to scenario fulfilment – a phenomenon that results in the neglect or distortion of information which is contradictory when faced with a stressful situation – which was a factor that came into play when the USS Vincennes downed the Iran Air flight. Robots are able to integrate larger volumes of information and sources far quicker than a human would be able to before deciding to use lethal force and they can be used to monitor human behaviour within teams of humans as a monitoring function to reduce the possibility of human ethical violations (Arkin 2009: 32).

In a later article, Arkin (2015: 46) reinforces the claim that humans have done a very poor job of being humane during battle and that we have a moral duty to protect the innocent non-combatants that continue to be victims in conflicts around the world. He believes that scientists have a duty to look to technology to reduce man's inhumanity. In fact, due to the low bar that has been set on the battlefield by humans, his claim that robots could one day be better at adhering to IHL is credible. If LAWS are judiciously designed and used properly, they could save non-combatants' lives and should be used to that end. It is not that he thinks that LAWS could be perfect ethical agents; he believes that this is not possible but that they could be more ethical than human soldiers (Arkin 2015: 46). The possibility that using LAWS could make combat more humane or ethical is an alluring possibility and some of the arguments raised are valid. Robots would not be subject to emotions or fatigue and could potentially process more information before making a decision. However, when considering the current level of technology, this might not be possible in the short term. It is also debateable whether this theoretical possibility should be used as a reason to pursue the development of autonomous weapons in the hope that it results in a more humane combat.

Generally, two classes of arguments present the use of LAWS as an ethical imperative. The first bases its pragmatic argument on the many documented failures in making lethal decisions in armed conflict and contends that probable or theoretical technological improvements and automation would result in better outcomes. The second argues is that these systems will reduce the risk to both combatants and, more importantly, non-combatants, as measured by a reduction in casualties. As such, it is morally necessary to use them. However, does a more precise weapon automatically make it more moral (Asaro 2012: 702)? These moral debates surrounding the motivation for the development of LAWS are controversial because they outline potentially valid and legitimate grounds for developing LAWS. However, these humane goals are currently potentially being dismissed given the humanitarian agenda put forward by the Campaign to Stop Killer Robots with the goal of banning these weapons on humane grounds.

5.5 COULD ROBOTS BE SUPERIOR ETHICAL ACTORS?

The idea that robots could prove to be superior ethical actors needs to be interrogated appropriately. When examining moral decision-making it usually falls into two approaches. The first is what is known as top-down morality where there is a precise application of rules that are used to make decisions by artificial agents. The second approach is that of bottom-up morality where there is an avoidance of referencing explicit moral theories by the development of systems that are able to tacitly learn to discern immoral and moral behaviour, a process which mimics human intuition. A hybrid approach can also be used as this combines elements from both these approaches (Bogosian 2017: 591-592). For a robot to derive its ethical system using a top-down approach, this is done with algorithms derived from rules for behaviour, or through a bottom-up approach where the rules are refined using experimental learning. When applying both approaches there needs to be a clear rule set from which to work. The added complexity is that if rules or goals are conflicting, the ability to effectively perform ethically will decline (Morkevicius 2014: 9). Therefore, the issue of having no clear rule set of ethical and moral behaviour results in difficulties for autonomous systems to implement a consistent and comprehensive ethical and moral framework.

Humans are able to give themselves moral commands. However, robots need to be given commands to follow. Machines are capable of replicating a moral action but morality does not emanate from following the actions of others. Therefore, a robot can act morally through mimicking actions designated as moral, but it cannot be intrinsically moral. An autonomous robot is a very complex machine that can interact with the world and respond to feedback. However, like a video of a moral act, the machine is just replicating the moral act. One would not say that the video is moral in and of itself, and like the video, the autonomous robot's actions cannot be considered to be moral as it is not capable of making a moral decision itself. This is because morals are based on values and values are made through sacrifices. Autonomous robots would be aware of their environment and could follow instructions. However, they are not aware of sacrifices and, therefore, robots do not have values of their own. As humans need values to think morally and give themselves moral commands, this can be thought of as a distinctive human trait (Johnson & Axinn 2013: 135). This

amounts to the idea that one can be moral by mere fact that one is human and that the ability to perform moral actions is not the determination of whether an agent and their actions are moral.

If a semblance of moral judgement is needed to make appropriate moral decisions and if LAWS are unable to replicate this moral judgement, then deploying LAWS to execute such decisions would be problematic from a moral perspective (Purves *et al*. 2015: 855). For this to be possible, robots would need to be programmed with the necessary ethical framework. However, this seems to be problematic for a number of reasons. For a robot to be a superior ethical decision maker, the robot would need to be able to make morally superior decisions more often than a human would. In the context of LAWS this would mean that the robot's ethical decision-making system would need to understand both international law and insights from thought on just war (Morkevicius 2014: 9).

An additional challenge exists in that, once a universal moral theory is agreed upon, the rules would need to be encoded. In order for clear rules to exist for LAWS to reference, it needs to be possible to codify moral principles. The codifiability thesis states that true moral theory can be reduced to rules that are universal and could then be competently applied by a morally ignorant person in every situation. However, the anti-codifiability thesis states that there is always the requirement that the agent exercise some moral judgement (Purves *et al*. 2015: 855-856). For Kant, morality requires that the principle of one's actions can be followed universally. This means that it is not the moral act that must be repeatable but rather the principle of the act. This ability to both judge and extract the principles cannot be codified and Kant believed that judgement could not be taught (Johnson & Axinn 2013: 135). McDowell (1979: 336) states that even if there was a genuine attempt to capture our perception of what virtue is and what it needed to be a set of rules, and if that process is considerate to capture all the subtleties and nuances in the resulting code, situations would still turn up where rote application of the code would seem wrong.

To illustrate this challenge, I will use the principle of distinction as required by IHL. The simple codification of the law is that, in battle, combatants need to be separated from non-combatants in order to only target combatants. This sounds easy if the battlefield

is open and all opponents wear a uniform that can be used as the basis for the targeting decision. However, this simplification ignores the added complexity of incapacitation or the act of surrender. Nevertheless, even with this simplistic scenario, it is possible to quickly blur moral logic. Early scholars who combined both secular legal reasoning with theologically derived Just War Theory stated that sometimes it is important to go further than what is legal in order to do what is right.

For example, what would a robot do if it came across an enemy combatant that had fallen asleep on the battlefield? There is no legal basis for the exclusion of this combatant based on international law. However, there is a history of Western soldiers struggling with this scenario as the sleeping soldier is not a direct threat as long as he remains asleep and killing him in his sleep somehow seems both morally wrong and unnecessary. An additional scenario is depicted in the film *Saving Private Ryan* where a battle inexperienced translator is frozen in fear in the midst of ongoing battle, both outside and inside the building where he is crouching and clutching ammunition on a flight of stairs. A German soldier walks past him after killing his fellow soldiers upstairs and pauses to stare at him. The translator is a uniformed, armed threat that legally meets the definition of a target. However, we instinctively know that if the German were to kill him, we would find it morally repugnant. Both these scenarios highlight that, even though they are both legitimate targets, in that moment in time they are not a threat.

Militaries would not want to punish soldiers who did kill these targets and in a counselling session might even remind them that they did their duty. However, the soldier who did not harm the individuals in these examples could also be praised, as long as it did not jeopardise the overall mission (Morkevicius 2014: 6). These scenarios highlight only two examples where killing would be legal but morally awkward, which poses a challenge for the codification of moral principles for LAWS.

These examples show that, even if it was possible to codify moral principles, they could in fact be so complex or context-dependent that it would render them incoherent. Furthermore, if they somehow were coherent, they would still need some form of moral judgement in the application of the code. Therefore, AI would never be able to replicate this moral judgment. Whatever intelligence it gains would result from a discrete list

provided by humans with no way of replicating human moral judgment (Purves *et al*. 2015: 857). This point was illustrated by Dreyfus (1992: 199) when looking at AI from a linguistics perspective. He states that programmed behaviour is either strictly rule-like or arbitrary. If a programme was given a new use for a phrase it would either have to treat it as a clear case which conforms to the rules of the language or it must take a random stab at what is meant. However, a native speaker has an additional option when confronted with the same phrase. She could recognise that the use of the phrase is strange and does not fit within the rules. However, she can understand to an extent what is being said and ascribe meaning to it within the context of human existence in what appears to be a very non-rule based, yet non-arbitrary approach. It is this intuitive understanding of the context that gives humans the edge over LAWS when interpreting a situation outside of the rules.

In addition to following moral and legal rules and principles, a soldier is also required to be able to disobey illegal orders. In order to be able to fulfil this requirement, soldiers would need to understand what illegal orders are and have the courage to disobey them. This requires moral judgement of a specific situation and, as robots have no values themselves, they have no basis for making such a moral assessment. In order to disobey illegal orders or violate regulations or laws, one would need to prove that the situation one faced was so unique that the violation of the law, regulation, or order was the moral thing to do. In other words, one would need to rely on human intuition which robots are unable to do at present (Johnson & Axinn 2013: 135). Again, here it is the human's ability to intuitively understand the situation without relying on a rules-based approach that makes the replication for LAWS problematic.

5.6 HUMAN DIGNITY

Human dignity is a pervasive notion that is embedded in national constitutions and IHRL. Some refer to human dignity as an empty space or deceptive *façade* primed for exploitation that threatens personal autonomy (Ulgen 2016: 3). But the introduction of the Universal Declaration on Human Rights (UDHR) states that inherent dignity stems from being human and that equal rights that are inalienable for all humanity is the basis of global peace, justice, and freedom. As the UDHR assigns all human beings inherent

dignity, there is the implicit assumption that all humans are worthy of respect (Docherty 2014). Dignity can be thought of as the property that makes humans eligible for the human rights listed in the UDHR (Johnson & Axinn 2013: 134). In fact, in IHL human dignity can be thought of as the conceptual underpinning and the *raison d'être* for IHL and IHRL. More recently, it has evolved into something so important as to have infiltrated the entire body of international law. The Martens Clause contains this elementary consideration of humanity and the ICJ has highlighted the Martens Clause as an effective tool to address the increasing technological advances in military technology. Human dignity can be conceptualised as follows: human beings have an intrinsic moral value from which flows certain rights, values, or duties implemented as rules related to how human beings should be treated in times of both war and peace (Ulgen 2016: 3-4).

Using Kant's moral theory on ethical conduct provides the basis for both the rationale and justification for these rules based on human dignity. In his search for a moral theory on human conduct, Kant postulated the idea of human dignity as a fundamental principle. This is captured first in *Groundwork of the Metaphysics of Morals* (1785) and later in *The Metaphysics of Morals* (1797). Both of these works demonstrate Kant's comprehension of the concept of human dignity as a privileged status awarded to humans from which undeniable rights and duties stem. This remarkable status is rooted in the capacity of humans to engage in rational thinking to create and abide by rules as well as the capacity to recognise "ends". In this context, Kant referred to "ends" as the reasons and justification for not only having rules but abiding by them. Linked to this is what is termed "autonomy of will" which refers to the capacity of humans to rationally and freely adopt and then abide by a set of rules. It is these features which provide humans with this overarching value that is human dignity. This dignity is unconditional with an intrinsic value which removes the dependency on any other factors for its respect, recognition, or existence. With this argument Kant elevates human dignity to a higher norm which allows humans to participate in the creation of laws and governance and gives them the ability to have certain moral expectations (Ulgen 2016: 4). In summary, according to Kant humans have dignity and should be treated as ends; objects have a price and can be used as a means to our ends. Robots, however, exist almost exclusively in a world of means mostly removed from ends (Johnson & Axinn 2013: 134). As a human you intrinsically

possess dignity which allows you to be a moral agent capable creating laws and governing. LAWS, on the other hand, do have dignity and their deployment could threaten human dignity.

Heyns (2017: 62) argues that dignity is what makes life significant because the right to life and the right to dignity are inseparable; both are interrelated and indivisible. This also applies to cases where there is a valid constraint on an individual's right to life. For example, when an individual is not protected by IHL, they still retain their other rights, including the right to dignity. In the case of LAWS, this dignity would be violated when an unthinking algorithm decides to kill an individual if they were a designated target. Such a death is indistinguishable from the deaths of others who find themselves to be victims of weapons that cause indiscriminate mass casualties such as WMD.

The central tenet of human dignity is that humans should not be reduced to something resembling an object, as in the case of rape or slavery, or something with no value at all, as in the case of victims of massacres. When LAWS direct lethal force towards a human, that human is reduced to an object to be demolished. This is exacerbated with the possibility of civilian causalities as there is no avenue to appeal to the humanity of the enemy when it is a machine that isn't capable of replicating humanity (Heyns 2017: 63-64). Rosert & Sauer (2019: 372) argue that, if the focus of the positions against LAWS is mostly on the indiscriminate nature of using LAWS, using them against only combatants would make them acceptable. However, combatants, too, have human dignity which should be respected. There is the additional challenge that when the arguments against LAWS have been focused on human dignity, it also focused on non-combatants, either civilians or soldiers who are in the process of surrendering or are *hors de combat*. Therefore, to boost the normative case against LAWS, it is necessary to explicitly state that allocating decisions about life and death to a machine is inhumane and unethical under any circumstance.

Rosert and Sauer (2019: 374) go further to argue that it does matter who kills a person. It matters to the person dying as they are reduced to data points which removes their right to be recognised as human in death. It also matters to the society which is responsible for the death. Contemporary warfare is already decoupling society from the financial and political cost of warfighting, especially in modern democracies. If that

same society now outsources the moral costs by not caring about the act of taking a life and without the burden and responsibility of killing, that society risks losing touch with both democratic and humanitarian norms. This further underscores the point that LAWS offer an avenue for the delegation of moral responsibility to machines, desensitising what it means to kill.

A machine's decision to kill would be based on a hypothetical set of pre-programmed instructions or rules of expected behaviour based on past experience. In the case of LAWS, the resulting decision to kill or not kill cannot be overruled as the human target has already been assigned a fate with no hope of altering the outcome (Amoroso *et al*. 2018: 32). This is similar to a mousetrap which kills targets based on characteristics and behaviour – meeting the target weight and either touching or eating the bait. The trigger of the mousetrap is intended to function based on the observation of the target and its actions while the intricacy of a mousetrap is of no concern. Because a mouse has limited dignity, it can be exterminated by a machine. The comparison can be made the LAWS are not enemy soldiers but in effect high-tech mousetraps with no concern for military honour and human dignity (Johnson & Axinn 2013: 134).

A real-world example of the violation of dignity that pre-programmed decisions bring about was brought before the German Constitutional Court which struck down legislation and ruled that the Minister of Defence could not order a hijacked passenger airplane to be shot down by the Luftwaffe despite confirmation that this passenger plane will be used to target people on the ground in circumstances similar to what happened in 9/11. The basis for this ruling was that human dignity is enshrined in Germany's basic law. In the act of shooting down the airplane, the passengers would be treated as objects and not people in achieving the goal of saving others on the ground. That would mean that, in achieving its goal, the state would violate their dignity by not recognising the worth of the passengers as human beings (Amoroso *et al*. 2018: 32-33). This illustrates Kant's argument that the passengers in this instance would be treated not as ends but as means to further other ends which violates the dignity bestowed on the passengers as humans.

A further point to ponder is that the dignity of those who direct the use of LAWS would also be compromised as dignity involves the capability to be a moral agent – this

means the ability to be autonomous, to exercise moral choices, and to assume responsibility for these choices. LAWS remove this possibility as the ability to be moral is outsourced to an amoral machine (Heyns 2017: 63-64). Therefore, the use of LAWS would seem to violate the dignity of all humans involved in or impacted by combat.

5.7 CHALLENGES TO HUMAN DIGNITY

The challenge to the ethical component of LAWS stems from the following question: what is so different from being instantly killed by a bullet to the skull or the heart fired by LAWS versus being killed by being set on fire or in the blast radius of a cruise missile? When signing up for the armed forces one understands the risks involved, including the potential to die. While the dignity argument resonates on an emotional level, it may just romanticise warfare. Human beings have conducted warfare on both a personal and industrialised scale for the last 200 years and it seems that there is little dignity attached to the firebombing of Tokyo or the 60,000 casualties on the first day alone that Britain suffered during the battle of the Somme (Horowitz 2016b: 32). The reality is that warfare is often ugly and inhumane by nature and that the deployment of LAWS would do little to worsen this situation.

Sharkey (2019: 84) argues that human dignity is not enough as a stand-alone position to argue against LAWS and should rather be clustered together with arguments about legal positions, technological maturity, and global security. This mainly stems from the absence of a globally recognised definition of human dignity. The dignity in question is also up for debate. Is it individual or collective human dignity that is under threat? There is also the question of whether human dignity is a human right in itself or is it the basis of human rights? Dignity also evolves over time and context. Therefore, dignity as a concept is not very concrete and thus provides a weak base to argue that LAWS as weapons are unique in challenging these ideas.

If one examines LAWS from a munitions perspective, it is unlikely that any significant challenge to human dignity outside of those already posed by current weapons will be presented. From a platform perspective, using LAWS as an anti-material weapon, for example against enemy ships in a clearly defined battlefield, is also unlikely to raise

any human dignity questions. However, in an urban environment or where the technological sophistication increases the risk of dehumanising targets, this becomes problematic. Human dignity also applies less when we are talking about LAWS that defend fixed installations. An analogy here is an electric fence. If the electric fence is clearly marked, an intrusion is considered a hostile action (Horowitz 2016b: 32-33). Therefore, LAWS as a category of weapons might not necessarily violate human dignity. Rather, it is the context in which LAWS are used that poses the challenge.

The biggest challenge to human dignity potentially comes in the form of LAWS being implemented at an operational system level where the conflict with Just War Theory is the most problematic. The main issue here is that, if LAWS were to control the decisions around whether and how to conduct military operations, this would be problematic as this would undoubtedly include handing off the moral responsibility to machines. However, if the systems were directing humans to choose and engage the targets in acting more like a high-level commander, this would pose less of a challenge (Horowitz 2016b: 32-33). Again, this highlights the complexity that where and how LAWS would be used would determine if they impact human dignity or not.

5.8 DESIGNING A MORAL MACHINE

A Moral Machine has been designed by the Scalable Cooperation at the MIT Lab. It is a platform which gathers human perspectives on the moral choices made by AI in applications such as self-driving cars. On the website users can offer their perspectives on which outcome is favourable when deciding on the circumstances presented – for example, whether it is better to kill five passengers or two pedestrians. Users are able to design scenarios and their inputs are also benchmarked against other responses (Scalable Cooperation 2021). When users are judging the different scenarios, they are presented with a number of variables to aid in their decision. The first is the different characters, their age, gender, social value, fitness, utilitarianism, and species. The second is that of their relationship to the vehicle as passengers or pedestrians. The third variable is the question around whether there should be intervention, i.e. no action being taken by the car or avoidance. The fourth is how the

rule of law impacts the decision, i.e. if the pedestrians crossed a red light (Verdiesen *et al*. 2018: 498).

Verdiesen *et al*. (2018: 501) have proposed a similar experiment for the development of a moral machine for LAWS and suggested a number of variables for inclusion into the design. These included the mission, location, type of weapon, number of characters, and the outcome. This could be a useful way of gathering insights about lay people's perspectives on the morality of LAWS. However, at this stage there is no evidence of the implementation of this moral machine on LAWS to date.

5.9 CONCLUSION

LAWS challenge the ethics and morality in war on a scale that has not been seen by a class of weapon before because of their autonomous nature. Life and death decisions would be performed by these machines and it is these decisions that are morally and ethically difficult. The first challenge is that there is no universal ethical and moral framework for LAWS to operate in. Even if this was in place, it might not be possible to reduce this to comprehensive and coherent rules for LAWS to operate within. This makes dealing with unusual but plausible battlefield scenarios difficult. LAWS also lack the moral agency to make these life and death decisions and can in some instances provide moral insulation for the human operators, as in the case of Iran Air flight 655. LAWS also challenge human dignity, even if the concept of human dignity has its own challenges. However, LAWS have also been championed as a potential tool to make warfare more ethical given that LAWS would not suffer from the same issues that make humans vulnerable to errors like anger or fatigue. The challenge that LAWS pose to the ethical and moral dimensions of war are perhaps the most problematic when considering their battlefield use.

6. CHAPTER SIX: CONCLUSION

6.1. INTRODUCTION

This chapter will summarise the key points that have been highlighted throughout this study. The research focused on LAWS as a new class of weapon or weapons system and how the development of this technology will challenge existing laws of war and the current understanding of ethics on the battlefield. It also highlighted the current attempts to regulate these weapons and the potential for a new arms race. This chapter will conclude with the key findings and suggest direction for future research.

6.2. RELEVANCE AND STRUCTURE OF THE STUDY

The objective of this study was to examine and analyse LAWS, focusing on their definition, their development, the challenges that they present to current interpretations of international law, the ethical issues they pose, as well as opportunities for regulation. Autonomous weapons pose a major challenge to the current international order as this is the first time that a revolutionary military technology is being developed almost exclusively in the private sector. This study intended to expand and add to current knowledge and broaden the understanding of the development of LAWS and the challenges that their development poses to current international law, ethical standards, and arms control regimes.

Chapter one outlined how the study and outlined how the study would be conducted. It provided an outline of the main challenges of LAWS and defined the research aim. It went on to provide a brief overview of the current literature on the topic and concluded with an explanation of the research methods used and the structure the study would take.

Chapter two began with defining AWS which led to a discussion about AI and how this technology makes LAWS possible. It introduced the concept of the OODA loop and outlined how this concept, when applied to battlefield engagement, allows for an explanation of how humans and autonomous systems can interact. The chapter described the different levels of autonomy and what the incentives are for adopting

autonomy for military applications. It explained that autonomous weapons already exist and discussed the first use of LAWS against humans. This led to a discussion about what could go wrong with autonomous systems, drawing on examples from the financial markets. It also discussed the future in terms of the possible uses of autonomous weapons as well as the potential for a new arms race and how LAWS could lower the threshold for conflict.

Chapter three began the analysis of the challenges LAWS pose to the current international legal framework, starting with Just War Theory and introducing the tradition of how warfare should be conducted. This progressed into a discussion of IHL and the two main principles found in this body of international law, namely the principles of distinction and proportionality, and the challenges the use of LAWS would pose to these two principles. It discussed the challenge of ensuring responsibility and accountability when deploying LAWS and how LAWS can be used to circumvent responsibility and accountability. Possibilities to overcome these legal challenges were discussed, including maintaining human control of LAWS or instituting Tort law to deal with transgressions. Finally, it summed up the relevant positions with regards to the whether IHL is sufficient in its current form to deal with LAWS.

Chapter four examined arms control and the incentives for adopting arms control for LAWS, including the possibility of a new AI arms race and the idea that subjecting LAWS to an arms control regime would reduce the security dilemma. It detailed the current framework of the CCW and the discussions that have taken place within the CCW as well as the guiding principles on LAWS that have been issued by the GGE. It discussed obligations that states have when assessing new weapons, including the obligation to conduct a new weapon review. The chapter then examined the role of norms and lessons that can be drawn from two previous successful attempts to regulate problematic weapons.

Chapter five discussed the ethical issues raised by LAWS. It explored the idea of a universal ethical framework and the need for moral agency when making life and death decisions. It also considered whether LAWS could potentially be more ethical than humans and the impact of LAWS on human dignity.

Chapter six concludes the study by providing a summary of the aim of the study and how the study was structured. It also summarises the key findings and suggests possible areas of future research.

6.3. OVERVIEW OF FINDINGS

Weaponised AI looks poised to change the future of how warfare is conducted around the globe forever. With the move from the theoretical construction and use of LAWS to the first documented deployment against human targets in Libya in 2020, it would seem that Pandora's box has been opened. While weapons capable of operating autonomously have existed and have been deployed for a number of decades already, this use of LAWS against humans has demonstrated that the future is already here.

There are debates about what constitutes AWS. It is clear from the study that this would constitute any weapon capable of identifying, tracking, and engaging a target autonomously without the need for instruction or consent from a human. With the advances in commercial AI and robotics for civilian purposes, there will be no way of regulating the development and transfer of these technological advances. The skill, knowledge, and resources required to adapt this civilian technology for military or terrorist purposes would be relatively simple and it is this ease of adaption that poses a threat to the stability of the global system.

While the CCW has had LAWS on its agenda for over seven years, it has failed to produce much other than guiding principles. While Germany, for example, believes that this is enough to be considered a legal instrument, Russia has no interest in concluding any legal instrument that would limit LAWS. The CCW relies on consensus and therefore getting to a binding agreement between the major powers within this process looks to be far off. Evidence from the ban on APL suggests that it may be possible to achieve a binding treaty amongst most members for limiting LAWS outside of the CCW. However, given the general availability of this technology, the effectiveness of this treaty would be questioned. An additional challenge to the effectiveness thereof is responding to the old idiom of fighting fire with fire. Will it be possible to respond to terrorist groups, for example, that are using even rudimentary

weapons capable of operating at machine speed? It seems unlikely that the states that possess the technology and capability to develop LAWS would forgo the strategic advantage these weapons systems can confer.

In conclusion, as these weapons are likely to become more common in the following years, the realms of international law and battlefield ethics and morals need to adapt to this new reality. Most of the debate up to now has focused on whether these two fields have advanced sufficiently to address this new challenge and it is pertinent to engage in new thinking about how these two fields would need to adapt their existing body of knowledge to incorporate the new reality that LAWS usher in. As very similar legal and ethical challenges would be faced in the civilian realm with regards to the development and use of this technology, this should provide a good starting point.

6.4. IMPLICATIONS FOR AFRICA AND AREAS FOR FUTURE RESEARCH

With the potential rise of LAWS there are twin security issues that would face Africa. The first, as witnessed in Libya, is the possibility of authoritarian governments to direct a cheap and effective means of suppression against their own populations. This capability could be used to entrench or seize power. The second threat is that posed by more technological advanced nations. As Africa is a net technological importer it is unlikely that there will be any real possibility of the development or manufacture of these types of weapons at scale in the near future. Therefore, African nations would be incredibly vulnerable to others who can deploy such weapon systems.

While this study has provided a comprehensive overview of some of the main issues when dealing with the issue of weaponised AI, various research themes have been identified that could be studied further in order to broaden the understanding that these weapons pose:

- What potential impact would the development of these weapon systems have on the security of African nations and their populations?
- The threat to human security in the light of nations deploying LAWS against their own populations.

- The impact that LAWS could have on enhancing the effectiveness of terrorist groups. Is the general availability of AI technology a boon for global terrorism?
- Flash wars, LAWS, and nuclear security. Is the interconnectedness of systems and the possibility of a flash war a potential trigger for a nuclear strike?
- AI and the future battlefield. Is the ability to perform at machine speed going to remove humans from battle?
- The prevention of an AI arms race. With rising tensions between the three great powers, is an AI arms race inevitable?

7. BIBLIOGRAPHY

Alexander, A. 2015. A Short History of International Humanitarian Law. *The European Journal of International Law*, 26 (1): 109-138.

Altmann, J. 2019. Autonomous Weapon Systems – Dangers and Need for an International Prohibition. In *KI 2019: Advances in Artificial Intelligence*, edited by C. Benzmüller & H. Stuckenschmidt. Cham: Springer Nature.

Altmann, J. & Sauer, F. 2017. Autonomous Weapon Systems and Strategic Stability. Survival Global Politics and Strategy, 59 (5): 117-142.

Amoroso, D., Sauer, F., Sharkey, N., Suchman, L. & Tamburrini, G. 2018. Autonomy in Weapon Systems: The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy. Heinrich Böll Stiftung. Internet: https://iris.unica.it/retrieve/handle/11584/247498/297745/Amoroso-Sauer-Suchman-Sharkey-Tamburrini_Autonomy%20in%20Weapon%20Systems%20%281%29.pdf. Access: 7 June 2021.

Arksey, H. & O'Malley, L. 2005. Scoping Studies: Towards a Methodological Framework. *International Journal of Social Research Methodology*, 8 (1): 19-32.

Article 36. 2016. Article 36 Reviews and Addressing Lethal Autonomous Weapons Systems: Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS). Internet: https://article36.org/wp-content/uploads/2016/04/LAWS-and-A36.pdf. Access: 9 August 2021.

Asaro, P. 2012. On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making. *International Review of the Red Cross*, 94 (886): 687-709.

Asaro, P. 2019. Algorithms of Violence: Critical Social Perspectives on Autonomous Weapons. *Social Research*, 86 (2): 537-555.

Bathaee, Y. 2018. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, 31 (2): 890-938.

Beard, J. M. 2014. Autonomous Weapons and Human Responsibilities. *Georgetown Journal of International Law*, 45 (3): 617-682.

Bode, I. & Huelss, H. 2018. Autonomous Weapons Systems and Changing Norms in International Relations. *Review of International Studies*, 44 (3): 393-413.

Bogosian, K. 2017. Implementation of Moral Uncertainty in Intelligent Machines. *Minds & Machines*, 27 (4): 591-608.

Boogaard, J. V. 2015. Proportionality and Autonomous Weapons Systems. *Journal of International Humanitarian Legal Studies*, 6 (2): 247-283.

Boulanin, V. & Verbruggen, M. 2017. Mapping the Development of Autonomy in Weapon Systems. Internet: https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf. Access: 21 October 2020.

Bull, H. N. 1961. *The Control of the Arms Race: Disarmament and Arms Control in the Missile Age*. New York: Praeger.

Campaign to Stop Killer Robots. 2019. About Us. Internet: https://www.stopkillerrobots.org/about/. Access: 20 May 2019.

Carrubba, C. J. 2005. Courts and Compliance in International Regulatory Regimes. *The Journal of Politics*, 67 (3): 669-689.

Chansoria, M. 2016. To ban or regulate autonomous weapons: An Indian response. *Bulletin of the Atomic Scientists*, 72 (2): 120-121.

Chengeta, T. 2016. Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law. *Denver Journal of International Law and Policy*, 45 (1): 1-50.

Convention on Certain Conventional Weapons (CCW). 2019. Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. Internet: https://undocs.org/pdf?symbol=en/CCW/MSP/2019/9. Access: 3 June 2021.

Coram, R. 2002. *Boyd, the Fighter Pilot Who Changed the Art of War*. New York: Hachette Book Group.

Creswell, J. W. 2014. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. 4th Edition. London: SAGE.

Crootof, R. 2014. The Killer Robots are Here: Legal and Policy Implications. *Cardozo Law Review*, 36 (5): 1837-1915.

Crootof, R. 2016. War Torts: Accountability for Autonomous Weapons. *University of Pennsylvania Law Review*, 164 (6): 1347-1402.

Dahlmann, A. & Dickow, M. 2019. Preventive Regulation of Autonomous Weapon Systems: Need for Action by Germany at Various Levels. Internet: https://www.swp-berlin.org/fileadmin/contents/products/research_papers/2019RP03_dnn_dkw.pdf. Access: 20 May 2019.

Docherty, B. 2014. Shaking the Foundations: The Human Rights Implications of Killer Robots. Internet: https://www.hrw.org/report/2014/05/12/shaking-foundations/human-rights-implications-killer-robots. Access: 6 June 2021.

Dreyfus, H. L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge Massachusetts: The MIT Press.

Eberlein, B. & Grande, E. 2005. Beyond Delegation: Transnational Regulatory Regimes and the EU Regulatory State. *Journal of European Public Policy*, 12 (1): 89-112.

Etzioni, A. & Etzioni, O. 2017. Pros and Cons of Autonomous Weapons Systems. *Military Review*, May-June 2017: 72-81.

Etzioni, O. 2018. Pros and Cons of Autonomous Weapons Systems. In *Happiness is the Wrong Metric a Liberal Communitarian Response to Populism*, edited by A. Etzioni. Washington: Springer.

Fereday, J. & Muir-Cochrane, E. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods*, 5 (1): 80-92.

Foy, J. 2014. Autonomous Weapon Systems: Taking the Human Out of International Humanitarian Law. *Dalhousie Journal of Legal Studies*, 23: 47-70.

Grant, M. J & Booth, A. 2009. A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies. *Health Information and Libraries Journal*, 26 (2): 91-108.

Grut, C. 2013. The Challenge to Autonomous Lethal Robotics to International Humanitarian Law. *Journal of Conflict and Security Law*, 18 (1): 5-23.

Guersenzvaig, A. 2018. Autonomous Weapon Systems, Failing the Principle of Discrimination. *IEEE Technology and Society Magazine*, 37 (1): 55-61.

Haas, M. C. & Fischer, S. C. 2017. The Evolution of Targeted Killing Practices: Autonomous Weapons, Future Conflict, and the International Order. *Contemporary Security Policy*, 38 (2): 281-306.

Haner, J. & Garcia, D. 2019. The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development. *Global Policy*, 10 (3): 331-337.

Hanson, M. 2012. Arms Control. In *An Introduction to International Relations.* 2nd Edition, edited by R. Devetak, J. George & S. Percy. Cambridge: Cambridge University Press.

Harford, T. 2012. High-Frequency Trading and the $440m Mistake. Internet: https://www.bbc.com/news/magazine-19214294. Access: 9 March 2021.

Hauptman, A. 2013. Autonomous Weapons and the Law of Armed Conflict. *Military Law Review*, 218 (2013): 170-195.

Hawkins, J. M. 2017. Thematic Analysis. In *The SAGE Encyclopaedia of Communication Research Methods*, edited by M. Allen. Thousand Oaks: Sage Publishing Company.

Heyns, C. 2016. Human Rights and the use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement. *Human Rights Quarterly*, 38 (2): 350-378.

Heyns, C. 2017. Autonomous weapons in armed conflict and the right to a dignified life: an African perspective. *South African Journal on Human Rights*, 33 (1): 46-71.

Hoffman, R. R., Cullen, T. M. & Hawley, J. K. 2016. The Myth and Costs of Autonomous Weapon Systems. *Bulletin of the Atomic Scientists*, 72 (4): 247-255.

Horowitz, M. C. 2016a. Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems. *Temple International & Comparative Law Journal*, 30 (1): 85-98.

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

Horowitz, M. C. 2019. When speed kills: Lethal Autonomous Weapon Systems, Deterrence and Stability. *Journal of Strategic Studies*, 42 (6): 764-788.

Human Rights Watch. 2012. Losing Humanity: The Case against Killer Robots. Internet: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots. Access: 20 May 2019.

Huntington, S. P. 1958. Arms Races: Prerequisites and Results. *Public Policy*, 8 (1): 41-86.

International Committee of the Red Cross (ICRC). 1899. Convention (II) with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land. The Hague, 29 July 1899. Internet: https://ihl-databases.icrc.org/ihl/INTRO/150. Access: 10 February 2021.

International Committee of the Red Cross (ICRC). 1977. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977 Basic Rule. Internet: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Article.xsp?action=openDocument&documentId=8A9E7E14C63C7F30C12563CD0051DC5C. Access: 10 February 2021.

International Committee of the Red Cross (ICRC). 1987a. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Commentary of 1987 Basic Rules. Internet: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=2F157A9C651F8B1DC12563CD0043256C. Access: 5 February 2021.

International Committee of the Red Cross (ICRC). 1987b. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Commentary of 1987 Duty of Commanders. Internet: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=36FC92EB9E83FBBEC12563CD00437BFB. Access: 3 March 2021.

International Committee of the Red Cross (ICRC). 2005a. Rule 14. Proportionality in Attack. Internet: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule14. Access 20 July 2021.

International Committee of the Red Cross (ICRC). 2005b. Rule 47. Attacks against Persons Hors de Combat. Internet: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule47. Access: 20 July 2021.

International Committee of the Red Cross (ICRC). 2006. A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977. Internet: https://shop.icrc.org/a-guide-to-the-legal-review-of-new-weapons-means-and-methods-of-warfare-pdf-en. Access: 4 February 2021.

International Committee of the Red Cross (ICRC). 2009. Protocols I and II Additional to the Geneva Conventions. Internet: https://www.icrc.org/en/doc/resources/documents/misc/additional-protocols-1977.htm. Access: 4 February 2021.

International Committee of the Red Cross (ICRC). 2010. War and International Humanitarian Law. Internet: https://www.icrc.org/en/doc/war-and-law/overview-war-and-law.htm. Access: 4 February 2021.

International Committee of the Red Cross (ICRC). 2013. Fully Autonomous Weapon Systems. Internet: https://www.icrc.org/en/doc/resources/documents/statement/2013/09-03-autonomous-weapons.htm. Access: 20 October 2020.

International Committee of the Red Cross (ICRC). 2015. What are jus ad bellum and jus in bello? Internet: https://www.icrc.org/en/document/what-are-jus-ad-bellum-and-jus-bello-0 Access: 20 July 2021.

International Committee of the Red Cross (ICRC). 2021. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Internet: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/States.xsp?xp_viewStates=XPages_NORMStatesParties&xp_treatySelected=470. Access: 8 February 2021.

International Military Tribunal Nuremberg. 1946. Nuremberg Trial Proceedings Volume 22: Two Hundred and Seventeenth Day Monday, 30 September 1946, Morning session. Internet: https://avalon.law.yale.edu/imt/09-30-46.asp. Access: 3 March 2021.

Irwin, S. 2013. Qualitative secondary data analysis: Ethics, Epistemology and Context. *Progress in Development Studies*, 13 (4): 295-306.

Jackson, R. & Sorensen, G. 2003. *Introduction to International Relations Theories and Approaches.* Second Edition. Oxford: Oxford University Press.

Jesson, J. K., Matheson, L. & Lacey, F. M. 2011. The Traditional Review. In *Doing Your Literature Review – Traditional and Systematic Approaches*, edited by J. K. Jesson, L. Matheson & F. M. Lacey. London: Sage Publishing Company.

Johnson, A. M. & Axinn, S. 2013. The Morality of Autonomous Robots. *Journal of Military Ethics*, 12 (2): 129-141.

Johnson, J. 2020. Artificial Intelligence, Drone Swarming and Escalation Risks in Future Warfare. *The RUSI Journal*, 165 (2): 26-36.

Kallenborn, Z. & Bleek, P. C. 2018. Swarming Destruction: Drone Swarms and Chemical, Biological, Radiological, and Nuclear Weapons. *The Nonproliferation Review*, 25 (5-6): 523-543.

Kirkpatrick, D. D., Fassihi, F. & Bergman, R. 2020. Killer Robot? Assassination of Iranian Scientist Feeds Conflicting Accounts. Internet:

https://www.nytimes.com/2020/12/02/world/middleeast/iran-assassination-nuclear-scientist.html. Access: 17 March 2021.

Klare, M. T. 2019. Autonomous Weapons Systems and the Laws of War. *Arms Control Today*, 49 (2): 6-12.

Kleinman, Z. 2020. Mohsen Fakhrizadeh: 'Machine-gun with AI' Used to Kill Iran Scientist. Internet: https://www.bbc.com/news/world-middle-east-55214359. Access: 17 March 2021.

Lachow, I. 2017. The Upside and Downside of Swarming Drones. *Bulletin of the Atomic Scientists*, 73 (2): 96-101.

Legg, S. & Hutter, M. 2007. Universal Intelligence: A Definition of Machine Intelligence. *Minds and Machines*, 17 (4): 391-444.

Le Moli, G. 2019. The Principle of Human Dignity in International Law. In *General Principles and the Coherence of International Law*, edited by M. Andenas *et al*. Leiden: Brill Nijhoff.

Leveringhaus, A. 2016. What's So Bad About Killer Robots? *Journal of Applied Philosophy*, 35 (2): 341-358.

Lewis, J. 2015. The Case for Regulating Fully Autonomous Weapons. *The Yale Law Journal,* 124 (4): 1309-1325.

Leys, N. 2018. Autonomous Weapon Systems and International Crises. *Strategic Studies Quarterly*, 12 (1): 48-73.

Liu, H. 2012. Categorization and Legality of Autonomous and Remote Weapons Systems. *International Review of the Red Cross*, 94 (886): 627-652.

Lucas, G. R. 2013. Legal and Ethical Precepts Governing Emerging Military Technologies: Research and Use. *Utah Law Review*, 2013 (5): 1271-1281.

Malik, S. 2018. Autonomous Weapon Systems: The Possibility and Probability of Accountability. *Wisconsin International Law Journal*, 35 (3): 609-642.

Marra, W. C. & McNeil, S. K. 2013. Understanding the loop: Regulating the next generation of war machines. *Harvard Journal of Law & Public Policy*, 36 (3): 1139-1186.

McDowell, J. 1979. Virtue and Reason. *Monist*, 62 (3): 331-350.

McFarland, T. & McCormack, T. 2014. Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes? *International Law Studies U.S Naval War College*, 90 (2014): 361-385.

Meier, M. W. 2016. Lethal Autonomous Weapons Systems (Laws): Conducting a Comprehensive Weapons Review. *Temple International and Comparative Law Journal*, 30 (1): 119-132.

Morkevicius, V. 2014. Tin Men: Ethics, Cybernetics and the Importance of Soul. *Journal of Military Ethics*, 13 (1): 3-19.

Mull, N. W. 2018. The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to it. *Houston Journal of International Law,* 40 (2): 462- 530.

Müller, V. C. 2016. Autonomous Killer robots are Probably Good News. In *Drones and Responsibility: Legal, Philosophical and Socio-technical perspectives on Remotely Controlled Weapons*, editied by E. Di Nucci & F. Santoni de Sio. New York: Routledge.

Newton, M. A. 2015. Back to the Future: Reflections on the Advent of Autonomous Weapons Systems. *Case Western Reserve Journal of International Law*, 47 (1): 5-23.

Noone, G. P. & Noone, D. C. 2015. The Debate Over Autonomous Weapon Systems. *Case Western Reserve Journal of International Law*, 47 (1): 25-35.

Noorman, M. & Johnson, D. G. 2014. Negotiating Autonomy and Responsibility in Military Robots. *Ethics and Information Technology*, 16 (1): 51-62.

Oxford Bibliographies. 2015. Jus Cogens. Internet: https://www.oxfordbibliographies.com/view/document/obo-9780199796953/obo-9780199796953-0124.xml. Access: 17 August 2021.

Persi Paoli, G., Vignard, K., Danks, D. & Meyer, P. 2020. Modernising Arms Control: Exploring Responses to the Use of AI in Military Decision-Making. Internet: https://unidir.org/publication/modernizing-arms-control. Access: 2 June 2021.

Petrenko, A. 2012. Between Berserksgang and the Autonomous Weapon System. *Public Affairs Quarterly*, 26 (2): 81-102.

Purves, D., Jenkins, R. & Strawser, B. J. 2015. Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice*, 18 (4): 851-872.

Ribeiro, M. T., Singh, S. & Guestrin, C. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Internet: https://arxiv.org/abs/1602.04938. Access: 13 January 2021.

Righetti, L., Pham, Q. C., Madhavan, R. & Chatila, R. 2018. Lethal Autonomous Weapon Systems [Ethical, Legal, and Societal Issues]. *IEEE Robotics & Automation Magazine*, 25 (1): 123-125.

Roberts, K., Dowell, A. & Nie, J. 2019. Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. *BMC Medical Research Methodology*, 19 (66): 1-8.

Robillard, M. 2018. No Such Thing as Killer Robots. *Journal of Applied Philosophy*, 35 (4): 705-717.

Roff, H. M. & Danks, D. 2018. "Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems. *Journal of Military Ethics*, 17 (1): 2-20.

Rosert, E. & Sauer, F. 2019. Prohibiting Autonomous Weapons: Put Human Dignity First. *Global Policy*, 10 (3): 370-374.

Rosert, E. & Sauer, F. 2021. How (Not) to Stop the Killer Robots: A Comparative Analysis of Humanitarian Disarmament Campaign Strategies. *Contemporary Security Policy*, 42 (1): 4-29.

Ruta, F. L. 2018. Do the benefits of artificial intelligence outweigh the risks? Internet: https://www.economist.com/open-future/2018/09/10/do-the-benefits-of-artificial-intelligence-outweigh-the-risks. Access: 10 September 2019.

Sassóli, M. 2014. Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified. *International Law Studies*, 90 (2014): 308-304.

Sauer, F. 2016. Stopping 'Killer Robots': Why Now Is the Time to Ban Autonomous Weapons Systems. *Arms Control Today*, 46 (8): 8-13.

Sauer, F. 2020. Stepping Back from the Brink: Why Multilateral Regulation of Autonomy in Weapons Systems is Difficult, Yet Imperative and Feasible. *International Review of the Red Cross*, 102 (913): 235-259.

Saxon, D. 2014. A Human Touch: Autonomous Weapons, Directive 3000.9 and the "Appropriate Levels of Judgment Over the Use of Force". *Georgetown Journal of International Affairs*, 15 (2): 100-109.

Scalable Cooperation. 2021. Moral Machine. Internet: https://www.moralmachine.net. Access: 10 June 2021.

Scharre, P. 2016. Autonomous Weapons and Operational Risk. Ethical Autonomy Project. 2016. Internet: https://www.files.ethz.ch/isn/196288/CNAS_Autonomous-weapons-operational-risk.pdf. Access: 9 March 2021.

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

Scharre, P. 2018b. Paul Scharre: Army of None: Autonomous Weapons and the Future of War. Internet: https://www.youtube.com/watch?v=Z9hVOCUUBvM. Access 14 June 2021.

Scharre, P. 2019. Killer Apps: The real Danger of an AI Arms Race. *Foreign Affairs*, 98 (3): 135-144.

Scharre, P. & Strategic Studies Quarterly. 2017. Highlighting Artificial Intelligence: An Interview with Paul Scharre Director, Technology and National Security Program Centre for a New American Security Conducted 26 September 2017. *Strategic Studies Quarterly*, 11 (4): 15-22.

Schmitt, M. N. & Thurnher, J. S. 2013. "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. *Harvard National Security Journal*, 4 (2): 231-281.

Sharkey, N. E. 2008. The Ethical Frontiers of Robotics. *Science*, 322 (5909): 1800-1801.

Sharkey, N. E. 2012. The Evitability of Autonomous Robot Warfare. *International Review of the Red Cross*, 94 (886): 787-799.

Sharkey, A. 2019. Autonomous Weapons Systems, Killer Robots and Human Dignity. *Ethics and Information Technology*, 21 (2): 75-87.

Sparrow, R. 2007. Killer Robots. *Journal of Applied Philosophy*, 24 (1) 62-77.

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

Steele, B, J. & Heinze, E, A. 2014. From Smart to Autonomous Weapons Confounding Territoriality and Moral Agency. In *The Future of Just War*, edited by C. E. Gentry & A. E. Eckert. Georgia: University of Georgia Press.

Surber, R. 2018. Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats. Internet: https://ict4peace.org/activities/artificial-intelligence-autonomous-technology-at-lethal-autonomous-weapons-systems-laws-and-peace-time-threats/. Access: 21 October 2020.

Szpak, A. 2019. Legality of Use and Challenges of New Technologies in Warfare – the Use of Autonomous Weapons in Contemporary or Future Wars. *European Review*, 28 (1): 118-131.

The Economist. 2020. Fighter Aircraft Will Soon Get AI Pilots. Internet: https://www.economist.com/science-and-technology/2020/11/19/fighter-aircraft-will-soon-get-ai-pilots. Access: 11 December 2020.

Thomas, B. T. 2015. Autonomous Weapons Systems: The Anatomy of Autonomy and the Legality of Lethality. *Houston Journal of International Law,* 37 (1): 235- 274.

Toscano, C. P. 2015. Friend of Humans: An Argument for Developing Autonomous Weapons Systems. *Journal of National Security Law and Policy*, 8 (1): 189-246.

Ulgen, O. 2016. Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an 'Elementary Consideration of Humanity'? *European Society of International Law Conference Paper Series,* 8 (9): 1-19.

Umbrello, S., Torres, P. & De Bellis, A. F. 2020. The Future of War: Could Lethal Autonomous Weapons Make Conflict More Ethical? *AI & Society*, 35 (1): 273-282.

United Nations Office for Disarmament Affairs (UNODA). 2021(a). High Contracting Parties and Signatories CCW. Internet: https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/high-contracting-parties-and-signatories-ccw/. Access: 1 June 2021.

United Nations Office for Disarmament Affairs (UNODA). 2021(b). Background on LAWS in the CCW. Internet: https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/. Access: 1 June 2021.

United Nations Security Council (UNSC). 2021. Letter Dated 8 March 2021 From the Panel of Experts on Libya Established Pursuant to Resolution 1973 (2011) addressed to the President of the Security Council. Internet: https://digitallibrary.un.org/record/3905159?ln=en. Access: 1 June 2021.

United States. 2012. Department of Defense Directive Number 3000.09. Internet: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf. Access: 20 October 2020.

United States Air Force. 2009. United States Air Force Unmanned Aircraft Systems Flight Plan 2009-2047. Internet: https://fas.org/irp/program/collect/uas_2009.pdf. Access: 22 October 2020.

Verdiesen, I., Dignum, V. & Rahwan, I. 2018. Design Requirements for a Moral Machine for Autonomous Weapons. In Computer Safety, Reliability, and Security. SAFECOMP 2018. Lecture Notes in Computer Science, vol 11094. Edited by B. Gallina, A. Skavhaug, E. Schoitsch, F. Bitsch. Springer, Cham.

Wagner, M. 2014. The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Weapon Systems. *Vanderbilt Journal of Transnational Law*, 47 (5): 1371-1424.

Wallach, W. 2017. Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles. *Communications of the ACM*, 60 (5): 28-34.

Wong, Y. H., Yurchak, J. M., Button, R. W., Frank, A., Larid, B., Osoba, O. A., Steeb, R., Harris, B. N. & Bae, S. J. 2020. Deterrence in the Age of Thinking Machines. Internet: https://www.rand.org/pubs/research_reports/RR2797.html. Access: 17 March 2021.