MAPPING THE FIELD OF AUTONOMY

A techno-anthropological study into the literature of autonomous weapons systems

AAU CPH Supervisor:

Søren Nors Nielsen

Email: Snn@bio.aau.dk

Master Thesis in Techno-Anthropology

Number of Characters / Normal Pages:

166,017 / 69

Student:

Alexander Nordal Behrndtz - (20153071)

Email: abehrn15@student.aau.dk

AALBORG UNIVERSITY
DENMARK

Dansk Resumé

De seneste års fremskridt indenfor autonome har revolutioneret den moderne slagmark og metoder til at føre krig. Alt peger på, at dødelige autonome våben vil blive den nye standard i aktive krigszoner i en nær fremtid. Udviklingen indenfor særligt informationsteknologierne driver denne udvikling fremad, og maskinlæring har løftet potentialet for kunstig intelligens til nye højder. Med sådanne muligheder, kommer der også store udfordringer. Konsekvenserne af disse teknologier vil ikke kun være moralske og etiske, men juridiske, politiske, fænomenologiske, sociale, psykologiske og meget mere. Men centralt i diskussionen er der opstået spørgsmål om ansvar. For hvem er ansvarlig for selvtænkende teknologier? Hvad sker der, når et teknologisk system opfører sig uforudsigeligt eller (i menneskeligt perspektiv) på ulogiske måder? Vil robotter en dag stå foran en domstol for deres handlinger, eller er ansvaret kun menneskeligt? - Dette er nogle af de spørgsmål, der stilles i debatten om autonome våbensystemer i dag. Denne afhandling er ikke et forsøg på at løse alle disse udfordringer, men er en eksplorativ rejse ind i felten omkring autonome våbensystemer. Ved hjælp af en aktør-netværks teoretisk og postfænomenologisk forståelse forsøger denne afhandling at fremme forståelsen for, hvordan denne emergerende teknologi medfører forhandlinger og translationer på kryds af store aktør-netværk, samtidig med at den forsøger at præsentere en teoretisk ramme for videre undersøgelser ind i dette komplicerede felt.

Acknowledgements

I owe many thanks to a great deal of people for helping me reach this point.

I would like to give thanks the Institute for Military Technology for my time as being a part of the institute as an intern, but also for supporting me and helping me during the process of writing this thesis. I would like to thank each and every member of the institute. In no particular order, I owe many thanks to Jeppe Teglskov Jacobsen, Andreas Immanuel Graae, Lasse Kronborg, Iben Yde, Katrine Nørgaard, Henrik Palbo, Niels-Thomas Rønne Ackermann, Søren Ankjer Strunge and Thomas Galasz Nielsen. Thank you all for clever discussions, framework, and support.

I would also like to thank my supervisor, Søren Nors Nielsen, who have guided and supported me through a tough and challenging process, and especially for his patience and flexibility with me. I would also like to thank my lecturers from Aalborg University, who have supported and gently yet sternly pushed me forward the past five years of my life.

I also owe a special thanks to my friend Ian Feole, for patiently enduring my long talks on the subjects of this thesis, providing feedback and support, and for uncomplainingly reading and correcting my grammatical and sometimes peculiar writing.

Finally, I would like to thank my family, especially my mother and my brother, as well as my close friends for all of the emotional support they have shown. Without their encouragement and aid, I would have given up long up long ago..

List of abbreviations

AWS	AUTONOMOUS WEAPONS SYSTEMS
ANT	Actor-Network Theory
(UN) CCW	Convention on Certain Conventional Weapons
HRW	Human Rights Watch
ICRAC	International Committee for Robot Arms Control
ICRC	International Committee of the Red Cross
IHL	International humanitarian law
IMT	Institute for Military Technology
LARS	Lethal Autonomous Robots
LAWS	Lethal autonomous weapons systems
LOAC	Laws of Armed Conflict
MCDC	(NATO) Multinational Capability Development Campaign
MHC	Menaingful Human Control
NAGSF	NATO Alliance and Ground Surveillance Force
NGO	Non-Governmental Organisation
PP	Postphenomenology
RDDC	Royal Danish Defence College
RPA	Remotely Piloted Aircraft
SDA-LOOP	Sense, Decide, Act-loop
STS	Science- and Technology Studies
SWAACS	Systems With Autonomous Attack Capabilites
UAV	Unmanned Aerial Vehicle
USDOD	US Department of Defense
UN	United Nations

Table of Contents

0. Preface (Motivation) – Who am *I* to Write This Thesis?

The goal and purpose of this thesis and my approach originates in several reasons and because of this, I feel that it is only fair that I reveal my motivation and framing to the reader, before embarking into this report in earnest. My motivation and framing of this report did not appear to me in the dead of night right before I had to finalize my master's thesis, rather, this thesis is the result of an ongoing literature study, that I have been undertaking both consciously and sub-consciously for about a year now, - perhaps even longer. My interest in military matters and technology is something that have followed me for longer than I can remember, if not to mention my relations and prior experiences with the Danish military. These important life experiences have left me with a keen interest in the workings of military technologies and organization.

During my childhood, I always had an interest in and connection to the Danish military. Much of my family were in or currently are in the Danish military. My father was a sailor in the Danish navy, when he met my mother, and so I grew up listening to many of his stories. My uncle on my mother's side, who I have spent many days admiring, has sailed all his life in service of the Danish navy, and he has also been a key figure in a line of many more military figures, whom I have grown up loving and admiring. As such, I cannot deny that I have always had an admiration of the Military, perhaps even a romantic perception on some parts. And as a young man at the age of twenty-one, I even volunteered for her Majesty's the Danish Royal Life Guard, an act which had many detrimental consequences on my life, but which is still something that I will never come to regret.

During my limited service, I had an accident which would impact me on a scale I could have never imagined beforehand. I was on an exercise, when my spine could not handle the weight of my gear and the physical stress, that I put on it. My vertebra dislocated and my life changed forever. I went from a physically healthy and fit soldier to a bedbound and broken individual. It is not difficult to say, that the time right after this accident was some of the hardest days of my life. Nevertheless, I had to acknowledge that my life was changed, and that a new reality gazed me in the eyes, one where I could no longer remain in the military, and where I would not get to live out the dream I had first imagined.

Perhaps some of the guardsman's "discipline" or "pride" was still left in me, or perhaps it was simply plain stubbornness, but I was not ready to just stay in that state. I went through therapy and rehabilitation, I trained and exercised every day,as I was determined to find a way back to life, and so I did. And this is the short story of how I ended up as a student of techno-anthropology at the university of Aalborg. While I was rehabilitating and receiving treatment, I realized, that I would have to find a new way in life, and so began my studies in the summer of 2015. And though I never lost my interest in the Military, its technologies, or its matters, I found new perspectives, thoughts, and interests in the world of academia, which would have otherwise eluded me. It is ironic to think, that without this fateful accident, I would probably never have found any interest in academia nor later been able to go on and gain an academic internship at the Royal Danish Defense College (RDDC). I would never have been able to combine these two worlds, that I find so interesting and which have inspired me in so many ways. The world of military and the world of academia. And even though my vertebrae are now somewhat back in their place more than five years after the accident, I never expected to find myself writing a master's thesis.

1. Introduction

The idea of automating warfare is an ever-expanding notion, perhaps as old as war itself. But as military technology advances at an accelerating rate, new and unexplored dilemmas emerge at an even faster pace. It was in February 2001, that the first Unmanned Aircraft Vehicle or UAV: RQ-1 Predator drone - was successfully armed and tested with Hellfire missiles. It was quickly deemed a success, and the RQ-1-predator not only shifted its attributes away from reconnaissance, it was modified in a way that changed its very function and essence. Replacing ***R****econnaissance* with ***M****ultipurpose*; the RQ-1 was transformed into the MQ-1, and at the same time marking an important transformation in modern warfare (Sharkey, 2009). As such, when the United States and its allies invaded Iraq in 2003, armed MQ-1 Predator drones began flying through the skies of the Middle East, supporting the war on terror. Though it was only a few at first, by 2010 these numbers had increased to several thousand unmanned aircraft, and though the type of drones were numerous and varied both in size and weaponry, several Predators, or the latter and perhaps even more famous model: 'MQ-9 Reaper'(s), were in the skies at all times (Cullen 2011, 17). Critics and sceptics of these war machines argue, that this transformation of the RQ-1 Predator into the MQ-1 should be regarded as a: *"historic revolution in warfare"* and that it should be viewed as: *"the most important weapons development since the atomic bomb*" (Singer 2010, as cited in: Cullen 2011, 17-18). As such, the advancement from the Predator to the Reaper heralds a new age; an age where autonomy and computer intelligence will soar, and responsibility for war actions will become ever harder to apportion (Cullen 2011, 19; Noorman & Johnson 2014, 51).

Since the evolution of industry, machines have become able to remove and eliminate tasks, which were inherently performed by humans. In our modern era, this includes war machines, and therefore also the Predator- and Reaper drones, however, even as the Reaper is capable of maintaining several tasks by itself, it is fundamentally reliant on the operator and pilot to choose, activate, and initiate its functions. It is not autonomous, though it can act and perform numerous actions on its own (Cullen 2011, 35-45 & 47-48). But, if the Reaper, and other drones like it, are not autonomous per se, why is there then a highly contested discussion on autonomous military technology surrounding the use of these drones?

As cars, phones, airplanes, household equipment and other technologies become “smarter” and incorporate increased intelligence as well as further autonomised systems, it becomes continually evident, that the future will be filled with systems capable of removing human involvement altogether. Intrinsically, military technologies are likely to follow this trend, - if not lead it. Progressively more intelligent and autonomous systems are potentially the fighters of tomorrow’s war. These systems might come in physical forms as machines, actively changing the physical world around them, but they can also come in the form of algorithms designed to control, flood, or stir friendly and/or enemy networks and change the flow and understanding of information (Scharre 2015, 3-4). As technologies incorporate new and greater levels of autonomy, the paradigm of controlled unmanned or remotely piloted technologies have the potential to quickly shift towards uncontrolled autonomous machines. It is not hard to imagine a future, where a pilotwill have the ability to oversee or control a “swarm” of military vehicles, - if the pilot is even needed at this point (Scharre 2015, 4). Although the Reaper-drone by itself only accounts for a fraction of the potential complications, which autonomous military technologies might present, it has become part of a representation of future military robotics. Even though it constitutes partially autonomous technology, it becomes a fact, that unmanned aerial vehicles are presently the driving force of military robotic systems, and as such, the Reaper, Predator, and other military drones easily becomes the center of discussion (Bieri & Dickow 2014).

While much of the focus has been with the Western Nations, especially American usage and development of autonomous technologies during their deployment in the Middle East, recent history calls for a change in focus as other powers of the world have been successfully attaining, developing, and using unmanned and/or autonomous technology. One need not look long at the world map to find military powers with capabilities either on par with or close to the United States of America. China, Russia, Israel, India, Turkey and many more (Arkin 2017, 38-39) have shown themselves as key players on this growing list of countries, which all utilise and develop unmanned and/or autonomous weapons systems successfully. This only serves to mark an even greater need for knowledge on the implications of these technologies, and even though the International Committee of the Red Cross (ICRC) raised concerns in 2011 about the ethical, legal, and societal challenges which autonomous weapons systems might produce (ICRC 2011, 36-42), debate is still fervently ongoing on the very fundamentals of the presented concerns. Scientific and public debate is unending, and as advances in robotics, sensors, computer systems and more keep leading to even more sophisticated autonomous technologies, new guidelines and doctrines are sought and explored (Yde 2016, 1; Nørgaard 2019, 21; Noorman & Johnson 2014, 51-52; Boulanin 2015). But, as these technological advancements have the potential of changing and reforming the very shape and understanding of warfare, this subject is under constant scrutinization, critique and evaluation. New Perspectives, arguments, and understanding are constantly being added within the field of autonomous weapons. As such, this thesis is not an attempt at discovering the "truth" about autonomous weapons technologies. This thesis is an attempt to examine and discern how some of these debates about autonomous technologies are understood, represented, and discussed. This is all done in an attempt to further the understanding of the consequences of what these technologies might bring, as well as how to face some of these effects.

1.1 From Autonomous Weapons to *Lethal* Responsibility

With the advances within autonomous technologies, it is no longer hard to believe, that Lethal Autonomous Weapons Systems (LAWS) could become a standard in active warzones in a near future, if not tomorrow. The introduction and evolvement of drone technology, machine-learning, improved A.I., and other contemporary information- and communication technologies (ICTs), the possibilities of war-machines have reached an all new high. But as should be evident by now, these possibilities have led to hard questions and impasses have emerged. The effects of these new technologies are not only moral, ethical, and technical, but also legal, phenomenological, social, psychological, political, and much more. The challenges emerging around these technologies can truly be labeled as *wicked*[1], with no single solution or simple answer available. Central to the discussion are the issues of responsibility and accountability. Who stands responsible when technology acts on its own? What happens if a technological system behaves and decides for itself in unpredictable or (in the eyes of humans) in illogical ways? Will robots one day stand themselves in front of a court for their actions or is responsibility inherently human? – These are some of the major questions being asked in the debate about autonomous weapons systems today (Noorman & Johnson 2014, 51-52).

[1] In this sense, wicked refers to the idea of a **wicked problem**, which is an often-used term used to describe problems that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize.

However, if we cannot hope to fathom all the potential consequences of these types of autonomous technologies and their effects, shouldn't we then simply abandon their usage? – This is a well-used (Sparrow 2007; Sauer 2016; and Surber 2018) and a fairly recognized and discussed argument (Thurnher 2015; and Geiss 2016), which might at first seem to be a straightforward solution, as if we merely legislate against any type of autonomous machine or system, we could perhaps avoid the wickedness of this problem altogether. The coalition "Campaign to STOP killer robots"[2] is a key example of the opposition that has formed in recent history as a result of the evolvement of these technologies. They are a community that includes many organisations, experts and movements, and have managed to gain various influence in the arena(s) surrounding autonomous weapons technologies with their goal of establishing a complete and indefinite ban on autonomous weapons[3]. But, as with many things political, legal, - if not to say human; the simple answer might not be so simple after all. Though many parties (mostly NGOs) are in fact pushing for state parties to negotiate and adopt a pre-emptive ban on autonomous weapons systems, only a few states have expressed a readiness for this discussion so far (Boulanin, 2015). Most states are still determining their positions on the matter. Though generally states wholly agree that the usage of Lethal Autonomous Weapons Systems that cannot comply with international law should be avoided, which also goes to restate the importance of the Geneva Conventions Article 36 of Additional protocol I (Boulanin, 2015)[4]. As such, this thesis is not an attempt at solving all of the problems described above, the goal is instead to firstly understand the emerging controversies and how they are represented across the scientific and political field, and as such rather an endeavor at sorting and exploring the "battlefield" consisting of definitions, discussions, interpretations, and understandings surrounding autonomous weapons systems. Subsequently, in this report I seek to add a Science- and Technology Studies lens to the discussion of autonomous weapons systems in an effort to provide a theoretical framework, that might help reveal unexplored avenues of handling the emerging controversies of autonomous weapons systems. I do this by combining an Actor-Network Theory framework with notions from Postphenomenology. Though these two theories come from two distinct traditions, which explore the world in fairly different manners, it is my belief, that a methodologically reasonable framework can be achieved. My hypothesis is that these two theories does not necessarily contradict each other, rather, I believe that when applied in the right manner, they can provide new unexplored understandings. In

[2] https://www.stopkillerrobots.org/about/#about
[3] For more information on their full program, see: https://www.stopkillerrobots.org/learn/#problem
[4] See also: United Nations' CCW report(s) "Convention on Certain Conventional Weapons" for more.

many cases, these two theories would create conflicting analytical perceptions, but it is my intention to explore, through a functional and pragmatic usage, whether an advantageous combination is feasible. As such, I attempt in this thesis to first apply an Actor-Network Theoretical lens to understand and investigate the relations and worlds of autonomous weapons systems. On the background of this analysis, I then apply a postphenomenological embodied analysis to better understand how these technological relations mediate human understanding and action.

2. Problem Area

Based on the background described above, this thesis seeks to explore how the socio-technical world(s) of autonomous weapons systems constitutes itself, as well as exploring the emerging controversies and how the different representations of the field influence these controversies. This study aims to accomplish this by applying an Actor-Network Theoretical and Postphenomenological combined approach, which will challenge the limits and potentials of both theories. As such, the following Research Question has been developed:

How does the socio-technical field of autonomous weapons systems constitute itself, and how does the constitution affect the perspective on the controversies of autonomous weapons systems?

To help answer this research question, I have constructed the following sub-questions:

1. How does the representations of autonomous weapons systems affect users of the technology?
2. Can the postphenomenological understanding be combined in a reasonable fashion with Actor-Network Theory to explain moral action and perception?

2.1 Autonomy in Action: When Autonomy Becomes Weaponized

But, before I start this process, I wish to outline some of the current popular represented understandings of autonomy in technology. Movies, documentaries, Science-Fiction and more have excitedly told stories of autonomous technologies which revolutionizes or transforms the very worlds in which they are presented (examples are: Terminator 1984-1991-2003 etc., Blade Runner 1982, 2001: a space odyssey 1968, and many more). And though these technologies might at first have seemed unattainable, the beginning of the 21st century has told a different story as autonomous technologies have truly started to see the light of day. The media and scientific communities seem to both eagerly carry this concept forward, despite the enduring debates and controversy revolving around this type of technology. Though many scientists and engineers differ much on their perspectives of the possibilities with autonomous technologies, various types and prototypes have emerged over the recent years.

While autonomous technologies come in many shapes and functions, with some of the most prominent being within logistics and transportation, other technological products have recently managed to gain unprecedented traction, popularity, and notoriety within both media and scientific communities. The star of these discussions has also managed to gather several names, which are being used almost interchangeably: Autonomous Weapons Systems (AWS), Systems With Autonomous Attack Capabilities (SWAACs), Lethal Autonomous Robots (LARs), and Lethal Autonomous Weapons Systems, (LAWS) or, as they are also often referred by sceptics and opponents: "Killer robots" [5], are an actual and very controversial topic presently, and not something which is only ascribed as part of 'science fiction' any longer. As such, politicians, NGO's, experts and more have formed many discussion forums on this subject.

Although, controversies surrounding evolvement of military technology is nothing new, and have often in the past as well as in the present been followed by numerous normative questions about the very nature, meaning, and consequence of these weapons (Nørgaard 2019, 20-21). These same questions are found in the discussions about autonomous weapons, but this debate seems almost as fervent when compared to

[5] That "success has many fathers", can also be said about the referring of these autonomous technologies. As such, I, in this report, view these terms as interchangeable, and though I will mostly use the term Autonomous Weapons System/s (AWS). I will at times also use other terms when this stimulates understanding or furthers the narrative.

many former or other current military technology discussions. Often being compared or analogized to that of the atom bomb or other doomsday scenario weaponry (see e.g.: Sparrow 2007; Singer 2010; Coeckelbergh 2013; Sauer 2016).

Katrine Nørgaard (2019) describes the debate on the utilisation of military robots as a battle of two technopolitical narratives, which both expresses vastly different perceptions on the future consequences of this technology (Nørgaard 2019, 20-21). These two narratives each try and shape the understanding of how these technologies will come to be used as well as what consequences might follow. On the one side, we find a narrative supporting military technological advances as a gateway into safer and more controlled warfare, where the risks of war will be minimized, as robots will replace humans on the battefield. Collateral damage will become rarer than ever before, combined with unprecedented situational awareness, control, and direct efficiency; warfare will become a completely different and more controlled setting. This is the narrative, which *technological optimism* attempts to advocate (Nørgaard 2019, 20). On the other side of this debate is a quite different story: With the dehumanization of warfare, through usage of autonomous military technology, we will not come to realize minimizing consequences and effects of armed conflict, rather, "killer robots" will lead to further destabilization, militarization as well as igniting arms races at an unprecedented level on a global scale. This narrative is instead labeled as *technological pessimism* (ibid.) and these two conflicting narratives can be found throughout much of the literature of AWS and other autonomous technologies. This depiction of two narratives or "camps", with vastly different opinions on AWS and other autonomous technologies, also appears to be something that goes beyond just one author's perception and is instead fairly accepted within the literature surrounding autonomous weapons technologies. It is one of the few normatives, which can be easily and formally identified relatively equally across scientific literature revolving around these technologies (see for more: Bieri & Dickow 2014; Geiss 2015, 3-6; Yde 2016 1-5; Scharre 2015; Williams 2015; Scharre 2016 6-8; Boulanin 2016; Arkin 2017; Bode & Huelss 2018 397-398; Graae 2019 9-12).

However, as Nørgaard also argues, both the narrative of technological- 'optimism' and 'pessimism' appears guided if not to say dictated by an underlaying principle of technological determinism, which causes these technologies to be represented as if the result of a natural linear progression or development. There is an internal logic that technology is always 'going forward', as well as a perception of technology as a regulating- or influencing factor to societal change (Nørgaard 2019, 13-21; see also: Graae 2019, 8-12 & 34-39; Noorman & Johnson 2014, 51-52), - and per se; co-regulating the very nature and/or understanding of warfare. As such, whether supporting the 'dystopian' or 'utopian' future, the deterministic belief produces a narrative where technological development is perceived as a "train which cannot be stopped" (Nørgaard 2019, 20).

Nonetheless, even if this technological train keeps going or not, the understanding of- and position on the very definition of autonomy is not homogenous, rather, different experts, states parties and organisations, such as the International Committee of the Red Cross, have different views on the meaning of autonomy (Boulanin, 2015; Williams 2015, 27). This entails, that even the boundary between automation and autonomy is an ongoing discussion as well, making it difficult to define a clear distinction based on policy or scientific definitions (Boulanin, 2015; Nørgaard 2019b, 99). In truth, even the very notion of autonomy as an attribute is exceedingly underspecified and hard to define. In humans autonomy is often viewed as going hand-in-hand with the concept of free will. Autonomy relates to the ability to carry out choices, make decisions, and being able to establish one's own rules and morality. 'How autonomy comes to be, what it actually is?' and 'can we even understand it in ourselves?' – are questions in a highly contested, protracted, and antiquated, yet ongoing philosophical debate across numerous fields (Noorman & Johnson 2014, 52-53). Making it no wonder as to why technological autonomy is an even trickier and obscure dilemma to comprehend. As such, it is also not the goal of this thesis to unravel the definition of autonomy and solving its underlaying dilemmas, instead this thesis focuses on exploring how the controversies around this term emerge in an STS-fashion.

3. Adding an STS-Theoretical Understanding

In the following, I wish to outline my theoretical understandings and the concepts which act as a foundation for my interpretation and investigation of autonomous weapons technologies. My theoretical understandings are constituted by several notions from Actor-Network theory as explained and defined by Bruno Latour and Michel Callon, and I combine this with key concepts and understandings from postphenomenology as presented by Don Ihde and Peter-Paul Verbeek.

3.1 Actor-Network Theory – a Language, a Lens, and a Method

My theoretical framework is, as mentioned, partly constituted by Actor-Network theory (hereinafter also referred to as ANT) which relates , as the name implies, to the study of actors and the networks in which these emerge. I do this mostly with the understandings and inspirations from Bruno Latour and Michel Callon; two of the key figures of ANT.

Actor-Network Theory originates from the science of sociology, and builds on a material-semiotic[6] understanding, in which human actors and/or non-human actants[7] are explored and researched through their relations in networks consisting of heterogenous relations. The networks consist of many varying entities, human as well as non-human, all connected in various degrees of stability, power, and magnitude (Jensen 2003, 2-7). As such, ANT provides a language capable of identifying and describing agency between humans and non-humans equally in this network of entities. This is a particularly useful characteristic, which is suitable for exploring autonomous technologies. ANT's practical use for "opening" up a field overflowing with various actants, both in terms of size and power, describing these and their

[6] Material-semiotics is applied as a framework within many different STS-genres. Though it differs in its use between traditions, it implies that relations between entities, which can be both human and non-human, should be perceived as equally and simultaneously material (it follows the things, and how they are shaped by their relations) and semiotic (how concepts and their relations and meaning are related). It assumes that there is no single social structure or form of patterning because these material and social networks come in different shapes and fashion. For more, see e.g.: http://www.heterogeneities.net/publications/Law2019MaterialSemiotics.pdf

[7] In the literature of ANT, "actants" are often used interchangeably with the human "actor". This is because of ANT's non distinction between human, object, or hybrid. As such, actant can be used to describe all of the entities equally without the need to distinguish. In this report, I also use actant interchangeably with the term for human actor.

relations is important in attempting at properly understanding and portraying the network surrounding autonomous technologies. As such, ANT was also my initial theory for exploring this field and how the various actants relate to each other.

Actor-Network Theory does not just focus or address one “world”, but rather concerns itself with multiple overlapping worlds, as such, relative size of an actant or even a network does not matter and can fluctuate between being microscopic to infinite. Across these “worlds”, in which the networks constitute themselves, all relations between actors and actants are perceived as *performed*, which relates to ANT’s semiotic origin. Everything is seen only in relation to something else, and they should and can only be understood in this fashion (Law 1999, 3-4). This is also one of Actor-Network Theory’s inherent flaws, as it can be a challenging feat to determine a single unanimous understanding or definition of ANT, and as a result ANT comes in numerous forms and has been used in many different ways (Jensen 2003, 4-9; Law 1999, 10-12).

In Actor-Network Theory, all entities are perceived as agents of change and mediation, and only through exploring their relations and by following and describing the networks, in which these emerge, can we gain an insight into the multiple heterogenous social worlds that exists around these. These networks then themselves consists of the same human and/or non-human entities, as well as hybrids hereof (Jensen 2003, 8-9). The world, as understood through the lens of ANT, is then an unstable one, as all entities are in constant battles, and where the victors are never unshakeable, as their very foundation builds upon their interconnectedness to each other. This is why Torben Elgaard Jensen (2003) expresses that ANT is not about reasonable negotiation, but rather that:

> “*Actor-Network Theory is on the bottom line not about fair tradeoffs, rather it is about parasitism. An actor-network is a chain of connections, where advantages are systematically distributed in one direction, so that a few actants enable themselves to speak for the many*” (Jensen 2003, 9).

I will return to this statement by Jensen, but firstly, I will expand on an important ontological understanding within ANT.

3.1.1 Generalized symmetry

Generalized symmetry is a notion conceived by Michel Callon, which embraces that both social- and natural sciences as equally disputable and ambiguous (Kroustrup & Olesen 2007, 73-74). It is an essential constructivist framework for understanding how phenomena should be viewed and understood in an ANT analysis. In Actor-Network Theory, all phenomena are perceived as in constant confrontation, as none of them are "truer" than the others. This means, that social and natural phenomena are not stable actualities, but are instead negotiable "truths", which help form our understandings and ordering of phenomena around us. Because of this, society cannot be regarded as a stable entity, rather, it must be viewed as unstable and unresolved (Kroustrup & Olsen 2007, 74). By following this notion, ANT creates an analysis open to controversies and interchanging understandings of what is social and natural, where nothing is perceived as certain or stable. This ANT analysis then bases itself upon identifying the relevant actants by describing their relations to each other, through observing how they engage, translate, negotiate, consolidate, and even combine in different ways, all in an attempt to map and understand what is happening.

An example of this is when Bruno Latour and Steve Woolgar (1979) conducted a study on the construction of scientific facts in their volume 'Laboratory Life'. In their work, they showed how scientists in a laboratory performed, negotiated and constructed different scientific facts, and Latour and Woolgar even went as far as to show, how the facts, that they presented in their volume were constructed in the same manner (Jensen 2003, 9-10). Latour and Woolgar discovered that facts were constantly negotiated by the researchers in the laboratory through different means, but often by using references, citations or using scientific instruments in scientific discussions (Latour & Woolgar, 1979, 86-87). Their detailed analysis of the laboratory showed that none of the negotiated facts were independent, constant, or certain. In short, this means that even in the scientific or socio-technical world, which Latour and Woolgar explored, facts were not stable, but instead negotiable. Central to this construction process is the focus on the idea of inscription devices, which act as a medium to understand and interpret reality (e.g. calculator or measurement tools). The inscription will then be transformed into expressions and meanings and are likewise applied in constructing facts (Jensen 2003, 10-11; Latour & Woolgar 1979, 45-ff).

In the same manner, in this report, I explore the negotiation and construction of facts surrounding autonomous weapons technologies. Though I am not in the laboratory with the scientists, I do see comparatives to Latour's and Woolgar's studies, and I find that mimicking their concepts and analysis in part can help me identify and understand the mobilisation of both human actors and non-human actants in this socio-technical scientific setting, that I've embarked upon. Actor-Network Theory offers a lens and language which allows for a deeper investigation of how the phenomenon of emerging autonomous weapons technologies affects the socio-technical world. But, to help use ANT to accomplish this, I also look to Michel Callon's notion of *Four Moments of Translation*

3.1.2 Callon's Four Moments of Translation

Translation as a notion origin from Michel Serres, who defines it as a form of mediation, which both furthers and interferes with a given signal, message or understanding. In this sense, translation can either help order or disorder given understandings, etc. (Jensen 2003, 8). Translation in terms of ANT is used to describe the way an actor or actant associates with other actors in the network. The actant gain strength through its association with other actants, and the strength becomes visible through its ability to set itself up as a spokesperson for the many (ibid.). As such, the translation process refers to the idea, that a connection is made between two or more actants, which were not there beforehand.

Jensen (2003) presents a few examples of successful translation processes as: a politician speaking on behalf of their electorate, a scientific report which states certain facts, and a wristwatch which shows time (Jensen 2003, 8). In all of these examples, the actants are only stabilized through their further affiliation with other actants. The clock is stabilized through a beforehand set of rules relating to the function of time and the measurement of same, which it then becomes the spokesperson for. The science report is stabilized through reference to other science reports, articles as well as the report's own inquiries, experiments, and results. Hidden in the network surrounding the politician is a landscape of election campaigns, ideology discussions and understandings, stakeholders, meetings with electors and more. As such, these examples can in themselves also lead to further actants, who are involved in different ways and who are also part of other networks and so forth.

Through exploring these relations between actants and describing their respected networks, ANT, understood in this way, provides a keen insight into the (in)stability of networks consisting of humans and non-humans alike (Jensen 2003, 8-9). This returns us to the fact, that all actants in ANT are agents of change. All actants: human as well as non-human - can employ a program or anti-program, which in turn results in negotiations, alliances, stability and/or instability. As such, processes of translations help stabilise the network(s) through extending statements, thereby folding the heterogeneous actors and actants, and redefining relational definitions between all entities in the respected networks. Latour defines statements as: "*anything that is thrown, sent or delegated by an enunciator*" (Latour 1991, 106). This is why, that through exploring the processes of translation, we may find a network's ordering effects (Law 1992, 386).

Michel Callon then further details this process by dividing it into what he describes as *four moments of translation*. The first moment is named *problematization* and involves how the actant or actants tries to dictate reality and its underlying problems in certain ways, as well as how these should be addressed (Callon 1986, 6; Jensen 2003, 18-19). Callon's second moment is called *interessement*, and it describes the moment, when an actant attempts to lock other actants into defined roles, which will in some manner assist the problematized programme. As such, interessement is about redefining the identities of the actants surrounding a network, removing their competitive connections to other problematizations of other networks or actants. It is about establishing devices or alliances, which can then separate the inter-*essed* actants and all other agents or entities who might define their identities in other ways, than what is wanted (Callon, 1986, 8-9). Callon describes it quite clearly: "*Interessement is the group of actions by which an entity […] attempts to impose and stabilize the identity of the other actors it defines through its problematization*." (Callon 1986, 8). Thirdly, are what Callon names the moment of *enrolment*, where actants negotiate, determine, and test their identities. Callon uses the neutral term of enrolment to encompass all types of actants, and then uses this term to describe the many ways a programme can make actants take on a specific wanted role (Jensen 2003, 20). This is done through different types of negotiations which might come in forms of tricks, seduction, physical violence or other.

> "*Enrolment does not imply, nor does it exclude, pre-established roles. It designates the device by which a set of interrelated roles is defined and attributed to actors who accept them. Interessement achieves enrolment if it is successful. To describe enrolment is thus to describe the group of multilateral negotiations, trials of strength and tricks that accompany the interessements and enable them to succeed*." (Callon 1986, 10)

As such, enrolment is dependent on interessement, but interessement is also only a success if enrolment is achieved. The fourth and last moment, which Callon defines, is the moment of *mobilization*, where a spokesperson (human or non-human) establishes itself as the voice of the many. If mobilization is successful, then an actant becomes a representative for a tacit mass of other actants in the network through its connections, aligning understandings and meanings with its own. Though, this should not be seen as an end result, but rather as a still negotiable role, which can be opposed at any given time and as such is neither stable nor given. (Callon 1986, 12-15; Jensen 2003, 20).

Finally, these four moments of translation, which Callon describes, additionally allow for an understanding of how Obligatory Passage Points (hereinafter also referred as OPP) emerges. OPP refers to access points, which must be passed in order to achieve a "goal", "need" or even sometimes simply understandings (Kroustrup & Olesen 2007, 78). An OPP can therefore also be interpreted as a core argument of a program, which can help align actants and their roles. As such, an OPP can be anything from ideologic, technological, human, or hybrid in nature. By following translation processes, we can then likewise follow an OPP, and better understand how these helps align actants towards a unified goal or understanding, thereby stabilising itself and the network surrounding it. Even so, translation processes should never be understood or perceived as final or complete, but rather be regarded as a continuous negotiated process, which will interchange between the different moments all the time (Kroustrup & Olesen 2007, 78, Callon 1986, 17-19).

As Callon himself describes it: "*Translation is a process before it is a result*" (Callon 1986, 19).

3.2 The Postphenomenological Turn: Ascribing Intentionality to Technology

The other half of my theoretical framework derives from Postphenomenology (hereinafter also referred as PP). Here I borrow most of my understandings from Don Ihde and Peter-Paul Verbeek, whom by many are regarded as pioneers within this theoretical setting. Postphenomenology has, with its inter-relational ontology and its understanding of humans, technologies and their connection to the world, an interesting view on the functions of the world and how it operates. In many ways, postphenomenology and Actor-Network Theory shares similar thoughts and frameworks. Both agree that human and world cannot be separated: "*When something in the world changes, so does the human change*" (Ihde 2015, iv). But they differ clearly in the philosophical traditions from which they have emerged. Where ANT draws its notions from semiotics, postphenomenology rather works in the world of embodiment analysis of human action and perception (Ihde 2015, iv), which leads to quite different potentials and limitations within the analysis and findings.

Postphenomenology draws its roots from phenomenological thinkers such as Husserl, Heidegger, Merleau-Ponty and others (Verbeek 2008). These early 20th century thinkers started to truly define 'classical' phenomenology as we know it today, working with understanding how humans and the world "truly" interact. And though these 20th century thinkers are often credited for defining phenomenology, it has come in many guises before throughout history. It is during the latest decades however, that a newer approach of working with phenomenology has emerged; namely postphenomenology. Largely because of Don Ihde's phenomenological analyses of human-technology relations, a new approach to the analyses of the social and cultural role of technology has surfaced (Rosenberger & Verbeek 2015, 9-13). Ihde began to use the term inter-relational ontology as a way of combining Heidegger's *dasein* or "being-in-the-world" and Husserl's intentionality. Through this, Ihde's focus shifted the very role of material technologies; he referred to these as mediating rather than fixed (Ihde 2015, xii). Ihde himself explains it as:

> "*Human-technology-world became the formalism expressing this inter-relationality. Humans actionally using technologies mediatingly relate to a world. What was to become my "phenomenology of technics" began here with a descriptive analysis of embodiment, hermeneutic and background relations.*" (Ihde 2015, xii)

Through this formalism, Ihde succeeded in creating a tool for analysis that exceeded where classical phenomenological approaches had fallen short. Classical phenomenology is criticised by many, including Rosenberger, Verbeek and Ihde, for its view of technologies as something which '*alienates*' humans from being *human*. This kind of analysis did bring technology into an important role of its time, as it allowed for an analysis of the social- and cultural role of technologies, but phenomenology began receiving significant criticism for being too 'romantic'. It lost touch with the actual roles, which technology had in regard to humans (Rosenberger & Verbeek 2015, 10). As such, postphenomenology distanced itself to be rid of this classical romanticism, while still analyzing and describing the world from a "close distance", something, they argue, that normative sciences cannot accomplish. Normative sciences will rather always look on a phenomenon from a "far" distance and as such only reveal a reduced world. Whereas postphenomenology takes us closer and "*takes us back to the things itself*" (Rosenberger & Verbeek 2015, 10-11).

> "*It is precisely this claim to regain access to an original world that is richer in meaning than the world of science and technology, that postphenomenology refutes. Rather than thinking in terms of alienation, it thinks in terms of mediation. Science and technology help to shape our relations to the world, rather than merely distancing us from it. This perspective of mediation embodies a reinterpretation of the foundations of phenomenology.*" (Rosenberger & Verbeek 2015, 11)

Even though I am not exactly investigating autonomous technologies at a "close physical distance" in this report, I still see many favourable analytical avenues by adding a postphenomenological lens as well. By first using an ANT-interpretation to understand the relations and "worlds" surrounding autonomous technologies, postphenomenology, with its focus on human-technological mediation, offers a theoretical framework which allows for a conceivable interesting and pragmatic analysis of how the technology might meditate the world to the individuals affected by this technology directly. By using ANT to firstly gain insights about the background and the translations and orderings surrounding autonomous technologies, I can then further add a postphenomenological analysis of how the world comes to be mediated and understood through the usage of AWS. Of course, this will only be in a limited fashion, but by combining these to theories in this manner, I believe a pragmatic and useful theoretical framework will come to fruition.

3.2.1 Technological mediation

By adding a postphenomenological understanding, technology is thus ascribed the role of *mediators* of human practices, morals, and choices. Not in the aspect that technology controls human action, but that it is through our interaction with technology, that new understandings arise which in turn leads to different and perhaps unexpected actions. Peter-Paul Verbeek (2006) elaborates, that technology should not be viewed as only functional or neutral instruments, but rather as mediators for human understanding and choice. This mediation happens through interaction with technology in specific contexts. As such, technology does not gain a role as leading or demanding human agency, rather it allows and *implies* various kinds of action. The technology also "disappears" from human attention pre se, as we instead get focused on the very practices, which the technology enables and facilitates. Instead of acting *with* the technology, humans are acting *through* technology and thereby creating their very relation to the world (mediated by the technology) (Verbeek 2006, 3). The technology becomes '*embodied*' by the user, and by acting through the technology, the world becomes *reshaped* or *mediated* (Rosenberger & Verbeek 2015, 14).

> *"Technological mediation is precisely this capacity of technology: technologies can mediate between humans and reality, by establishing specific relations between both. This*

phenomenon of technological mediation has two dimensions, each of them pertaining to one aspect of the relations between humans and reality. First technologies help to shape how reality can be present for human beings, by mediating human perception and interpretation; second, technologies help to shape how humans are present in reality, by mediating human action and practices. The first dimension can be called hermeneutic, since it concerns meaning and interpretation; the second is pragmatic, since it concerns human activities." (Verbeek 2006, 3)

It is through this distinction quoted above, that Verbeek then splits mediation into two distinct dimensions, the *hermeneutic* and the *pragmatic*. This distinction allows for an analytic perspective into different users' mediated perception and interpretation (the hermeneutic dimension) of the world. Whilst also allowing for an analysis of the mediated effect which technology has on users' agency, action, and practice (the pragmatic dimension).

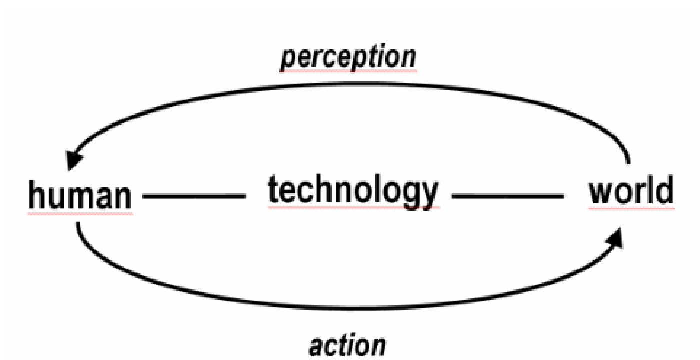

FIGURE 1. (Verbeek 2006, 3)

Expanding on Ihde's human-technology-world analysis, Verbeek added his two dimensions and named them *perception* and *action*, in turn allowing for an even further understanding of humans, technology and their interconnectedness to the world. It is through this framework that Verbeek goes on to talk about technological intentionality, which describes how the technology mediates certain actions or understandings. Verbeek relates this to what Latour and Akrich defines as '*scripts'* (Verbeek 2006a, 4; See also Latour 1992, and Akrich 1992). The technology becomes a script for human action and instructs users how to use it and/or how to interpret it. An example of this are speedbumps, - they cause drivers to slow down their vehicles for safety reasons. As such, speedbumps mediate through two different aspects.

Firstly, they make drivers perform an action: 'slow down to ensure safety', - this is the pragmatic mediation. But this action is only a result of an interpretation or *perception* of the world. The speedbump changes our perspective of the road and our possibilities. Drivers interpret the speedbump as a sign of: 'danger is ahead, if I do not slow down', - this is where the hermeneutic mediation can be identified. This is the essence of technological intentionality. Technologies change our action and perception of reality through hermeneutic and/or pragmatic mediation.

But in postphenomenology, technology does not always have a single way in which it is used or even conceived. In fact, it is the exact opposite. Technology in itself can be perceived and used in a variety of ways. Technology is something which we design and use for own goals and reasons but is also conceived as non-neutral and determining for human action and perception. This is the reason why Ihde created the notion of multistability in the 1980's (Rosenberger & Verbeek 2015, 25-26; See also: Ihde, 1986 or 2012). Multistability refers to the concept, that technology can have multiple purposes and interpretations. The technology might have been developed for specific situations or causes, but instead ends up getting used in totally unexpected ways (Verbeek 2006, 5). These unintended uses of a technology might lead to many different scenarios. The technology might mediate very differently than originally intended , though it need not be either wholly in a positive or negative way. Verbeek explains this with the usage of two examples:

> "*The energy-saving lightbulb is a good example here, having actually resulted in an increased energy consumption since such bulbs often appear to be used in places previously left unlit, such as in the garden or on the façade, thereby canceling out their economizing effect (Steg 1999; Weegink 1996). Second, unintentional, and unexpected forms of mediation can arise also when technologies do get used in the way their designers intended but meet unforeseen use practices. A good example is the revolving door which keeps out not only cold air but also wheelchair users.*" (Verbeek 2006)

These two examples from Verbeek really illustrates how technology can lead to unintended actions and/or perceptions regarding technology. To better understand how technological intentionality might occur, postphenomenology attempts to understand the relations between human and the technology which mediates the world. To accomplish this, Ihde made an addition to his human-technology-world and named this framework “human-technology relations”. To make a successful postphenomenological analysis, you must understand the human-technology relations. In fact, this is where many postphenomenological studies begin (Rosenberger & Verbeek 2015, 14; See also Ihde 1990).

> “*In order to understand a technology or a technological development, postphenomenology always analyzes the character of the relation human beings have with this technology and the ways in which it organizes relations between human beings and the world. Human beings can interact with technologies, incorporate them, read them. All of these relations organize how human beings experience their environment, and how they are practically engaged with it. Technologies, to be short, are not opposed to human existence; they are its very medium.“* (Rosenberger & Verbeek 2015, 12-13)

The human-technology relations that Ihde presents are divided into four types: embodied relation, hermeneutic relation, alterity relation and background relation. Though it is also argued that these four types of relations of mediation should not be seen as exhaustive, instead they are regarded as an analytical viewpoint from which to understand how users develop connections and usage of a technology (Rosenberger & Verbeek 2015, 14-19). But, this is also where I diverge from Ihde’s preconstructed human-technology relations, as I instead rely on ANT to interpret the connection and relation between human and technology. As such, though I adopt Ihde’s and Verbeek’s notions on how humans are mediated by technology pragmatically and hermeneutically, I instead apply ANT as a means to uncover how and of what these relations are constructed (See section 6).

But rather than utilizing the human-technology relations as presented by Ihde, I instead build on top my ANT-analysis to define the relation between the users and the technology. While I do view Ihde’s four predefined relations as durable, I do not view them as flexible enough to truly characterize the full scale of the relational context and its important significance. While I do not reject PP’s embodied understanding that intentionality is carried through the usage of technology, the postphenomenological analysis does not by itself give enough credit or focus to other underlying social and political effects, which affects the technological understanding, usage, and embodiment of the technology. ANT exceeds at explaining and outlining the constituted performed worlds surrounding AWS and how their interconnectedness constantly changes and shapes the field of autonomous weapons systems, I believe that postphenomenology is better equipped at outlining the users' mediated actions in a more pragmatic fashion. When operators and pilots activate or use remotely piloted aircraft, they perceive the world through this technology and its functions. The technology influences the understanding and thereby the actions; by means of mediation (See section 3.2). Examples can be borrowed from the medical world, where practitioners utilize numerous and varied technological tools to gain insight into different patients' conditions. A radiologist gains access to a patient's illness by using an x-ray scanner, a doctor assumes a patient's condition based on blood-samples and so on. In these cases, it is the technology that allows for interpretation of the patient's world and thereby also action. As such, these technologies are highly moral tools, which have a huge impact on users' choices and actions (Verbeek 2011, 1-11).

> *"When Technology are used, they always influence the context in which they fulfill their function. Technological artefacts help shape human action and perception, and create new practices and ways of living… When technologies fulfill their functions, they also help to shape the actions and experiences of their users. This phenomenon is called* ***technological mediation****: Technologies help shape the experiences and practices of their users."* (Verbeek 2011b, 201).

But where Verbeek (2011) argues that technology *only* act as moral mediators, and that technology cannot by itself have moral agency (Verbeek 2011, 52-53), as opposed to ANT, where all actants are perceived as equal agents of change, I rather utilize a dualistic framework, where I agree with both interpretations. This might seem contradicting at first, but as I will come to present, there is an advantage to be gained by accepting both notions.

By combining these two theories, I believe a framework can be established, where firstly an Actor-Network-oriented focus on how the different heterogenous networks of actants are performed, translated, and constructed can be applied. By following how the different human and non-human agents emerge across the interconnected worlds, insights can be gathered on how and *why* the debate around autonomous technologies is as it is. But where ANT exceeds at outlining and depicting why, where and how the socio-technical world(s) is and became, I do not believe that it gives enough explanation on the world(s)' *doing*. Therefore, I instead consider towards postphenomenology. By applying a postphenomenological analysis on the background of the ANT-uncoverings, it is my belief, that a more practical dimension can be attained, which opens an avenue for further investigation into the interaction between technology and human.

4. Methods and Reflections

In the next section, I elaborate on my methodological choices, usage, understanding and reflections. The purpose of this is to explain my methodological framework, and why I have chosen to approach my master's thesis in this manner.

4.1 Setting the Scene: overcoming bodily limits and the corona virus – How to write a master's thesis while bedbound

As mentioned in the preface, the questions asked in this thesis, did not just occur to me out of the blue, rather, they have been slowly developed over the last year. I became more and more involved and interested in the field of autonomous weapons systems during my time as an intern for the Royal Danish Defence College's (RDDC) Institute for Military Technology (IMT). I began this internship at the Institute for Military Technology in September 2019 with the goal of helping to solve and explore the many emerging complications which relate to autonomous technologies.

Originally, my focus was set on assisting in resolving issues which related to the weapons screening of new autonomous weapons systems. As weapon technologies and systems keep advancing, new models and systems are both sought and needed to review these weapons and understand their potential risks as well as proper use in accordance with Article 36 of Additional Protocol I of the Geneva Conventions. My main role as a techno-anthropologist was then to share and employ my methods and theories to help identify any relevant actors/actants and explore their involvement and relation in terms of said weapons screening. It was during this process, that I realised how complex the controversy surrounding autonomous weapons systems truly is. As I worked with on subject, read papers, and had conversations with researchers at RDDC on this field, my understandings deepened, and my interests peeked. But, the more I read and talked about AWS, the clearer it became to me, that this field is seemingly surrounded in shroud of mystery.

My original assignment for the Institute for Military Technology ended in January 2020, but as IMT was fairly pleased with my work, it was agreed to prolong my internship until the end of June 2020, as that would give me access to several texts, people, and more, which were meant to be, and at the time was, invaluable assistance for me to write my master's thesis. My plan became to use my contacts and network at the Royal Danish Defence College to gain actual physical access to the field of drones in Denmark, or at least gain access to those who occupy themselves with these on a daily basis. But, regrettably, my physical health started to fail during late December 2019, and it became apparent to me in February 2020, that my body and health had become a huge limitation. I became practically confined to my home, and I could hardly even manage to turn up at my workstation at IMT. A few weeks absence quickly became a month of sick absence, and I finally recognized, that I probably would not come back to IMT any time soon. I contacted Lieutenant Colonel Thomas Galasz Nielsen, the chief at IMT, to let him know of my situation; he promptly understood my condition and made it clear that I would be allowed to continue my internship from home.

Though definitely not a perfect situation, I still kept in contact with IMT, and I visited a few times to discuss possible routes and ideas for my project. It was at these meetings and "visits", that I got access to or was presented with considerable parts of the literature presented in this thesis. The IMT is based on the idea of gathering expertise from several different fields into one unit and making use of interdisciplinary cooperation, as such, the literature also turned out to be quite varied, and from many different scientific fields' point of view. This I believe, opened my eyes to the underlaying heterogeneity surrounding autonomous weapons technologies. I quickly understood that even the very idea of what is or isn't "autonomous" was a very discussed item and that this was in no way a formalized definition. As such, these discussions and talks I had early in this project's lifetime at IMT proved vital, as I realised, that I would need to explore several different perspectives.

But just as I was getting prepared to change my study focus from field observations into one of interviews, COVID-19 made its way to Europe. As if to kick me, while I was laying down, isolation and quarantine hit Denmark and the rest of the world like a hammer. Society went into lockdown and that included the Royal Danish Defence College and therefore also the Institute for Military Technology. All employees were sent home, to ensure that none would spread the virus, and only vital or secret work could continue at RDDC. Meetings and information were eventually moved online, but it took some time to organise, which meant that I was stuck at home with little to no information on the situation. It also did not help, that because of my illness, I was additionally told to avoid all physical contact with other people by the doctors treating me. This led to even further isolation and frustration. As even before corona, I had felt somewhat cut off from RDDC and even my university, but afterwards it became hard to stay in touch with any of my contacts or my network. Consequently, I had no choice but to accept, that I had to change my direction for the third time, and needless to say, this was a frustrating revelation.

How does one write a master's thesis, when both field, subject and informants are suddenly removed? At first, my answer was simple and exasperated, you do not. I was ready to completely abandon this project by March 2020, but after some reflection, I realised, that I *had actually managed to meet the field*. I had spent more than half a year surrounded by it. Though I had not interviewed any experts formally, I had instead *informally* interviewed several experts at RDDC during my time as an intern. These conversations or 'informal interviews' had already shaped much of my project's direction, as it was through these, that I ascertained the immense controversy surrounding it. And though I didn't have any notes or recordings of these 'informal interviews', they had led me to new insights and literature, which all told a story about the controversies surrounding AWS; some texts even written by the very experts, that I had wished to interview. These works of literature could then serve as first-hand narratives and descriptions, which I had full access to, even as I laid bedbound in my home. Each paper and book would serve as a direct representation of the field, that I wished to embark and understand.

4.2 Conducting a Qualitative Systematic Literature Review

The purpose of a literature review is to ensure a foundational and solid comprehension of the field of study. It allows for new insights to be discovered through synthesizing the findings of other sources into one (Green et al. 2006, 102). By gathering, building, and discussing earlier works within a field, the literature review has the attributes needed to design a space which is excellent for discussing philosophical discussions in a balanced manner (Green et al. 2006, 103). After my conversations at IMT, I had managed to gather a fairly substantial amount of literature, but, to ensure that I had a reasonable and fair foundation to build assumptions upon, I could not only rely on literature presented to me or otherwise gathered through my work at IMT. I would still need to conduct a literature review of my own to ensure a fair and representational understanding of AWS, and so I did. However, the time spent with IMT had supplied me with considerable amounts of relevant inquires to build my search upon.

By conducting a systematic literature review, a foundation can be built, which the rest of the report can develop and lean upon. While firstly illuminating known discussions and findings within the field, it also helps revealing lacks or gaps in the current body of literature (Green et al. 2006, 102). For this reason, I will firstly present my methodological considerations and means of conducting my literature review. Hereafter, I present and discuss the key discoveries gained through this manner, and lastly, I collect and sort the discussions while providing my own interpretation on the topics presented. To ensure proper direction in my literature search, I utilised the following key question: *What are the ethical consequences and moral discussions revealed and represented as in the research literature, and how are these presented and referred?* This question, along with my findings from IMT helped focus my search. As the field of literature about (lethal/) autonomous weapons technologies/systems is constantly expanding, - and includes works from many different fields of science, this key question was a crucial factor for ensuring a limited direction and focus. But, to ensure some room for surprises and to attempt overcoming bias, some search words intentionally deviated for the sake of ensuring a fair broadness in the literature review.

4.2.1 The Literature Search

I employed my gathered knowledge from IMT and focused on the complications, topics, and some selected key authors which I had encountered during my internship. These, together with my key question, served as inspiration for my literature search. Many of the authors, which I discovered through IMT, have also been included in my final literature review. I utilised three search databases for my literature search: Google Scholar, Aalborg University's online library: aub.aau.dk, and SCOPUS. Google Scholar served as a first-hand, broad, and extensive search engine, whereas aub.aau and SCOPUS allowed for more narrow and precise searches, where I frequently employed the function to narrow the search further by adding additional search words or specific authors' names. The choice of search words was not selected beforehand but was rather a result of my talks at RDDC, IMT, my research question, my key question, and my overall exploration of autonomous weapons technologies. This resulted in search words such as:

“Cyber AND autonomy”

“Autonomy AND STS-studies”

“Autonomous technologies AND actor-network theory”

“Ethics AND LAWS”*

“Autonomy AND NATO AND Ethics”

“LAWS AND the Geneva Conventions”*

“Autonomous AND weapons AND technologies AND Drones”

“Cyber AND autonomous weapons”

*“RPA** AND UAV AND studies”*

“Social understandings of L/AWS”*

“LAWS AND design”*

“Predator studies”

“Design AND Autonomous Weapons Technologies”

“Weapons screening AND L/AWS”*

“Design AND drones AND UAV”

“Drones AND STS”

“Lethal Autonomous Weapons Technologies AND NATO AND Geneva conventions”

“NATO AND weapons design”

“Reaper studies”

“Ethics AND drones”

“Drone technology AND policy”

and more.

*LAWS would in this case be written one as both LAWS and then as Lethal Autonomous Weapons Systems.

*AWS would in this case be written one as both AWS and then as Autonomous Weapons Systems

**RPA would also be written as both RPA and Remotely piloted Aircraft.

The results from my Google Scholar searches easily resulted in thousands, if not hundreds of thousands of hits, and as I did not have neither the time nor resources to go through it all, I instead decided that I had to accept Google’s search algorithms[8], and assume that the most relevant articles appeared within the first few pages. As such, I preemptively decided to only go as far as page five. As each page contained 10 search results, going to page five meant that I would sort through 50 results within each search.

[8] Google is known to use PageRank system, which “ranks” pages in its searches. PageRank works by counting the number and quality of links to a webpage to determine how important the website is. The assumption is then, that more important websites are likely to receive more links from other websites. See for more: https://web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglesearchworks.html

Search Words	Database	Search Hits	Chosen Article(s)	Preliminary notes	Publisher / Link	Other notes
Autonomous technologies AND actor-network theory	Google Scholar	15.800	None	No articles of interest		Remove or change search word
Autonomous Weapons Technologies AND actor-network theory	Google Scholar	4.440	2 hits (1) "Autonomous weapons systems and changing norms in international relations" (2) "A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous	(1) Discusses the effect of AWS and the way, these shapes or change norms through their practice. Discuss public legitimacy and expectations (2) An overview of ANT which also utilises examples of autonomous technologies. Might	(1) Cambridge core: Review of International Studies[1] (2) Simon Fraser University[2]	

Figure 2. – Example of usage of literature matrix (see appendix A.2 for full version).

I noted any relevant article, paper, book, or other text which I found. My requirements for the preliminary notation was, that the text had to include at least a substantial part about autonomous technologies. Then I also checked if the text was written by authors, whom I had discovered before, and/or if the text's keywords matched my key question. If any of these requirements were fulfilled, I would then go on further to read abstracts and/or introductions (and sometimes other parts such as: conclusions), which would give me an initial understanding of the perspective on the topic of the text. Depending on this preliminary read, I would then promptly decide whether the text was relevant or not to my further investigation. If it were relevant, I would add it to my *literature matrix*. To help organize my searches, I had adopted the idea of a literature matrix. This method allowed me clear oversight and transparency of the collected literature, as well as a solid platform to conduct my searches in an orderly and organised fashion (See figure 2. shown above for example).

The matrix was organised to include: 'used search words', 'search engine', 'amounts of search results', 'chosen articles', any possible 'preliminary notes', 'publisher's name', and 'optional or additional notes'. I added all initially relevant articles to my matrix, which resulted in a list containing 103 different articles, books, and other texts revolving the ethical- and moral discussions surrounding autonomous weapons systems. Though some of the texts were added for their perspective on other autonomous technologies, as these frequently followed many of the same dilemmas or questions represented in the discussion of AWS.

4.2.2 Saturation – Where does the Literature Review Stop?

To ensure, that I had "caught" appropriate amounts of literature, I employed a *capture-mark-recapture* method inspired by Kastner et al. (2007). The method is originally inspired by ecology studies, where fish from a pond, lake, or river is first caught, then marked, and finally released back into the wild. This allows the researchers a fairly representable estimate of the overall fish population through statistical analysis of the amount of captured and recaptured fish (Kastner et al. 2007). I, while not fully adopting Kastner et al.'s method, used this approach in the same manner. When I "caught" an article, I would "mark" it in the literature matrix, and I would then start to notice, if the same articles reappeared in other searches. This was especially easy when using Google Scholar, as it highlights all previous visited links. If the same articles started to reappear multiple times, it indicated that a satisfying level of saturation was being reached. Likewise, if already caught articles, books or other appeared in the reference lists of the selected text, I again saw this as a sign of saturation. This, together with my subjective reading of the articles' relevance helped me to assume when I had reached an agreeable amount of literature. I similarly utilised the same approach while reading the texts and conducting my literature review to assist me in knowing when I had gained a prominent understanding of specific topics, controversies or other. If I reached a point, where I no longer gained any significantly new or surprising insights through reading, I would then decide that I was "recapturing" insights, and therefore the amount of literature on the topic, that I was reading, would be representative of the field.

I then decided to restrict myself by means of relevance, and as such began by reading texts that I had recaptured more than three times. I began with the text with the most entries from my assembled table of collected literature (appendix A.1), and I then simply read downwards from the table. This left me with a total of 24 texts of different extent and proportion. These first 24 texts would then, together with my collected literature from IMT, serve as my platform for understanding the field. Some texts were discarded while reading, as I either found them uninteresting for my research- and/or key question, or they did not provide any useful information. Likewise, if they did not provide any new insights, but rather echoed previously discovered insights, I would note this and move on to the next text.

An important factor to mention here is that, while I read, I made sure to note the authors and their potential connections to different organisations or states. This was important, as autonomous technologies are a highly political subject, and as such, I always kept a note of who the text was made by and for. In example, the text I recaptured the most, was published by the 'ICT for Peace Foundation'. The ICT for Peace Foundation is an NGO which engages itself in ensuring that Information- and Communication Technologies are developed and used for peaceful and "*dignified*" purposes. They promote cybersecurity and peaceful cyberspace through engaging in international negotiations with governments and other actors (Surber 2018, ii). As such, they can easily be seen as having a motive and/or bias when releasing papers on autonomous weapons technologies[9]. But, as I am in fact interested in the very representation of the field, these possible biases do not make the texts less relevant, rather their representations are exactly what should be sought after. The important and challenging part in this was rather to ensure representation across all of the field, something that I had to subjectively attempt to accomplish.

4.2.3 Coding my Literature Review

To help organise the insights of my literature review, I used an analytical coding method. As a result of my preliminary research and inquiries from IMT, I managed to code the texts in relation to six predefined categories, which helped create an overview of the literature in regard to my areas of interest and key concepts. The categories were: (1) Descriptions and definitions of autonomy, (2) Understanding and/or representation of practice and use of autonomy in technology, (3) Types of technological understandings, (4) Details of- and types of technology presented, (5) Methodology used to examine autonomy, and finally (6) Ethical consequences and moral discussions; how they are revealed and represented. Furthermore, I also noted the presented controversies in the writings and how these were represented, as well as any notable concepts, which could help me define a further theoretical foundation for use in this report.

[9] See: https://ict4peace.org/activities/ for more.

Though this was somewhat limited, because autonomous weapons systems are researched within so many different research fields, which makes it difficult for me to include various concepts presented within or in collaboration with my own theoretical framework.

4.2.4 Interviewing the Books, Papers and More

Though it is impossible to interview or ask a book questions directly, there are still some advantages to have, by deliberately writing questions for books and "asking" the book or text to answer these. Following my six predefined categories presented above, in an attempt to help myself in exploring these categories, I produced several questions, which each category might provide answers for. Though these questions should not be regarded as exhaustive in all ways, nor as fully representational of the matters which I sought or found answers for. Instead, they were created as a tool, which allowed- and helped me to stay focused throughout my reading and exploration of the writings. These questions helped me define the categories in a functional manner, so that my coding could be achieved pragmatically and transparently. Though, if I did not encounter answers to my questions at all, I would reflect on why these questions had not been part of the paper or book, which aided me in both identifying *lack* or *abundancy* in the literature. Meaning, that I noted whether matters, concepts, or other reoccurred multiple times or perhaps only once or twice, or maybe not at all.

In the end, I arrived at the following table:

Theme	Goal	Research Questions	Follow-up questions
Descriptions and definitions of autonomy	To get an understanding of the different approaches to the term "autonomy", which seems to be understood fairly differently.	What is autonomy? Where do you "find" autonomy? What is autonomy defined as?	• What is the background of autonomy? • What is the difference between automation and autonomy? • Where do you see autonomy in action?

Understanding and/or representation of practice and use of autonomy in technology	To get a deeper understanding of the term autonomous weapons system in practice. When is it used, what does it entail, and how "deep" does this term go?	When is something termed as "autonomous", when is it "lethal" what is the difference of LAWS, SWAAC, and AWS? What does it mean to be a "weapons system"? Can sub-systems in themselves be termed autonomous?	• When is an autonomous weapons system lethal? • What is the difference between AWS, SWAAC, and LAWS? ○ Is there a difference? • How are autonomous weapons understood differently from other autonomous technologies? • How is autonomy in technology represented in practice?
Types of technological understandings	To gain a deeper insight into the technological understandings of the actors in the field. How are they viewed, understood, and more? What are the factors and who does the actors/actants see as involved here?	When are technologies unmanned or manned? What is the difference of saying RPA and e.g. UAV? How do drone-pilots view the technology? How does the public view the technology? How does designers view the technology? How does these groups view the consequences of this technology?	• How do autonomous technologies function? • What are the practices involved with using RPA as compared to UAV, or AWS? • How does the pilots view use of autonomy? • How does revolving parties view use of autonomy? • How are they made/designed and engineered?

Details of- and types of technology presented	To understand what is viewed as AWS or autonomous technology in general. It is important in order to better understand the classifications of the texts.	What kind of technologies are viewed and named as autonomous? Are there "good" and "evil" autonomous technologies? When does an autonomous technology become an autonomous weapons system/technology?	• Which technologies are usually labelled as autonomous? • What comes first, weapons technology or civilian? • Can cyber operations be called autonomous? • Are there any autonomous technologies being developed and/or used right now? • Why/why not?
Methodology used to examine autonomy	To understand how and through which lenses autonomous technologies are examined. I am especially interested in any STS-oriented methods.	What methods are used in order to understand or examine autonomous technologies? Are any methods viewed as "better" or "worse"? How big is the field of STS on autonomous technologies? And which methods/theories emerge?	• Which method(s) is/are used? • How are they used? Observation, review, other? • Are any methods seemingly better than others?
Ethical consequences and moral discussions; how they are revealed and represented	Ethics are an important part of technological development, perhaps more so with military technology. Therefore, my goal is to see how the literature presents	How does autonomy and other Information and Communications Technologies change the ethics of warfare? Have there been any change in ethics, or	• What have changed in war since the emergence of RPA/autonomy and more? • Has it become easier/harder to fight "cleanly"?

	understandings on ethics, - if it does!	do the same “codes” still apply? How are ethics generally viewed in in the text?	• How does actors/actants in the field view ethics? • Are ethics viewed as an important part of the discussion? – Why/ why not?

Figure 3. - Table of questions used doing literature review

Each category’s questions were divided into ‘goal’, ‘research questions’, and ‘follow-up questions’. I started by defining and outlining my goal for exploring the selected category, which helped producing various research questions for exploring each category in broad terms; usually resulting in questions such as: ‘What is autonomy?’ or ‘What does it mean to be a weapons system?’. These questions were then reduced or directed into more focused follow-up questions, but the border between these two should not be seen as static. This method of sorting my inquiries in this manner served as a means of sorting the texts in a practical way. I also utilised this method as a reflective tool, making me aware of my own pre-understandings and possible bias. By outlining my goals, research question and follow-up questions in this way, I also outlined sections of my subjective bias, which is especially important in understanding why I discovered, what I did.

5. Patterns of Autonomy – Mapping the Definitions

In this section I would like to discuss some of the insights I have gained through reading and coding the texts. One particularly important element of this literature review was the definition of autonomy. What is autonomy defined and revealed as? The ever-expanding role of autonomous systems maintains its momentum, and as such, challenges occur that has not before seen the light of day. One of the biggest challenges is ironically to pinpoint what the very term *autonomy* entails. As a result, this highly debated subject emerges as a paramount component in the texts explored. I therefore sought to investigate whether the current literature focus and consider one specific definition or several definitions, as well as how much this term influences the debate. As such, writings which had a substantial focus on the definition of autonomy were further considered. Hence, during my literature study, I discovered that some texts viewed autonomy as an all-encompassing term, whereas others viewed it as an almost inequitable term, - both in a variety of ways; while others either completely ignored or simply avoided this term and its underlaying debates altogether. One reoccurring theme, which many authors agreed upon, was that there is no consensus or single definition on autonomy, and many even questioned whether reaching a consensus would be possible. Andrew Williams (2015) describes how the adhesion surrounding the term of autonomy is leading to serious difficulties, as the debate is littered with conflicting and incompatible definitions of what autonomy should be understood as:

> *"The subject of 'autonomy' or 'autonomous systems' is of growing importance in both military and civilian domains. Technology is evolving rapidly and recent military operations increasingly rely on robotic and unmanned systems, which incorporate increasingly 'autonomous' functions. Multiple definitions and understandings currently exist about autonomous systems, and the related concepts of autonomy, automation, robotics, and unmanned systems. This terminological confusion makes basic coherence of defence programmes challenging, and at worst, affects strategic level defence policy when certain concepts are misrepresented in public debate."* (Williams 2015, 27)

But, even though there is no singular or homogenous definition of autonomy, there were some distinguishable themes that emerged during my literature review. More than a few define autonomy through "degrees" or "levels" (Grut 2013; Bieri & Dickow 2014; Geiss 2015; Bode & Huelss 2018. See also: Scharre 2015, 8-11). Whereas others rather focus on the distinction of autonomy through ethics. Often describing autonomy parallel with being able to conduct and/or understand moral actions or decisions (Arkin 2010; Englert et al. 2014; Geiss 2015; Arkin 2017; Surber 2018). While the former two groups of experts perceive autonomy from very to fairly distinct from automation, there are also those who argue, that autonomy in machines is relatively unattainable and/or that autonomy is specifically human and thereby argue, that you cannot differentiate between autonomy and automation in machines. They state that machines are all inherently automatic and will always only respond to their preprogramming, no matter how well programmed (Meier 2016; Umbrello et al. 2019; (Surber 2018); see also: Williams 2015, 52-53). But, even as I strive to thematize some of the emergent definitions, it should be noted that the boundaries which separates these definitions easily blurs and are hardly static. Nor is this attempt to thematize even exhaustive of all the different views on autonomy, instead it includes some of the larger and commonly represented understandings. Many different experts criticize, augment and/or combine interpretations from other experts and create new understandings or descriptions in new text, often restating in the conclusions the importance of gaining further insight into a communal definition of autonomy, - which seems almost poetically just out of reach every time.

The authors, who define autonomy in degrees or levels, usually argue along the lines that autonomy can be measured from "low" to "high" intelligence in terms of decision-making or task fulfillment, and typically relates how perceiving autonomy in this fashion can help sort legal and moral inquiries. Matthias Bieri and Marcel Dickow (2014), who work within the perception, that no matter the level of autonomy, a human operator should always be to some extent part of the procedure, defines autonomy in three varying degrees: (1) '*Remote Control',* where the user or operator uses the machine to conduct various tasks distant to the operator while the machine additionally supports the operator through various complexity reduction, such as translating sensor data into sensible information. (2) *'Autonomous Maneuvers under Human Steering Control'*, where the machine can autonomously perform its task, - such as flying and surveying, but the operator has the ability to constantly change and adjust its function or task. (3) '*Autonomous Execution of Tasks Without Human Control, but With Veto Right'* where a human operator can only stop or intervene in the machine's operations through a type of veto command. As such, the machine autonomously performs all its actions, and the operator's only action is to stop the machine in case of emergencies. (Bieri & Dickow 2014, 2). Essentially Bieri and Dickow does not perceive the human as completely replaced, but they raise concerns to whether further technological development will lead to humans being gradually removed from the "chain of decision-making and responsibility". Furthermore, they focus on whether an operator would be able to act in real time with the autonomous machine properly, and how to deal with this dilemma (Bieri & Dickow 2014, 2-3):

> *"The key question is whether the human operator is really aware of all that is going on in a given situation. The operator essentially perceives the world through the eyes of the assistance system. When experienced in real time, the reasons why an algorithm has taken a decision to act, or has preselected and suggested such a course of action, can no longer be comprehended. This problem has been existing for years."* (Bieri & Dickow 2014, 3).

As such, Bieri and Dickow's questions revolves around the issue of whether an autonomous machine, even under the eye of an operator, may possibly cause unforeseeable consequences, such as breaking obligations under International Law. Similarly, but not the same, Ingvild Bode and Hendrik Huelss (2018) definition of autonomous weapons systems equates to its degree of autonomy or "independent agency" (Bode & Huelss 2018, 397). Rather than defining clear or distinguishable levels, they work within a spectrum of varying degrees, where the simplest reactive mechanisms comprises one end, and human intelligence the other end. Because of this, they view AWS as something that is progressively changing towards human intelligence, and intrinsically defines AWS following this reasoning:

> *In this sense, AWS are defined as 'systems that, once activated, can track, identify and attack targets with violent force without further human intervention'. Hence, AWS qualitatively surpass remote-controlled systems, such as drones, that are currently important security technologies* (Bode & Huelss 2018, 397)

Essentially, Bode and Huelss maintains that AWS differentiate from other autonomous technologies, because of their growing "non-human agency", which is a result of the fundamental extent of decision-making which these technologies will need to be capable of if fully implemented. When instead reading texts from experts who define autonomy through a more moral or ethical driven argument, I found that the perspective also changes quite a bit. The understanding of "fully autonomous" machines and/or the idea of perception of levels or degrees of autonomy are often rejected or contested (Surber 2018, 5-8). Regina Surber (2018) instead calls for a distinction between human autonomy and a machine's ability for self-governance and/or decision-making. In fact, Surber argues, that the discussion on the term autonomy in relation to machines is misguided, as the word "autonomy" is just borrowed from philosophy's classical usage of the term, involving human's ability of self-governance and understanding moral responsibility. By using this same term for machines, we instead risk limiting human responsibility and relinquishing human accountability for potentially immoral or illegal actions; in the case of AWS: killing civilians or destroying civilian property (Surber 2018, 6-8).

Ronald Arkin is another referenced author in the field of autonomous technology and his perceptions seems to be representative of many of the advocates of autonomous technologies. Though Arkin (2017) does acknowledge, that adding autonomy to weapons systems or other machines complicates matters, he argues that this is more of a philosophical problem, rather than a functional problem (Arkin 2017, 35). Arkin argues that autonomy only involves "[...] *(the) delegation of decision-making to a machine that has been pre-programmed by a human."* (Arkin 2017, 35). As such, Arkin views other delineating discussions about autonomy such as the discussion on levels or degrees of autonomy as confounding and misleading for the potentials of autonomous weapons systems (Arkin 2017, 36). Through a series of moral arguments, Arkin argues that autonomous unmanned systems could be made morally superior to humans. The history on human behavior in warzones, Arkin argues, demonstrates that humans are inherently bad at adhering to legal and ethical requirements (Arkin 2010, 3). So instead, war would see many benefits from shifting the flawed moral human with the imperfect autonomous unmanned system. By illustrating a list of controversial military actions committed during the recent Iraqi-war, Arkin illustrates that only by enforcing Laws of Armed Conflict and international humanitarian law (IHL) in a new manner, can humanity come closer to a more ethical battlefield. This is where unmanned autonomous systems become essential:

> "*The primary goal remains to enforce international humanitarian law (or the Laws of Armed Conflict (LOAC)) in the battlefield in a manner that is believed achievable, by creating a class of robots that not only comply with the restrictions of international law, but in fact outperform human soldiers in their ethical capacity under comparable*

circumstances. If successful this will result in the saving of noncombatant life and property, ideally without erosion of mission performance. It is too early to tell whether this venture will be successful. (Arkin 2010, 10)

Essentially, Arkin argues that it is unrealistic to assume that human soldiers would ever be able to completely adhere to IHL or the Laws of Warfare, when they first confront the horrors of battle, and that with the increasing use of new weapons technology in the hands of humans, increasing terror and progressively worse atrocities will come to transpire. Instead humans should strive to restrain themselves and the human-technological warfighter, and this is where the case for ethical autonomy in unmanned systems reveals itself (Arkin 2010, 3-10). If we instead add the same logic to autonomous weapons systems, as we do that of the soldier under orders, then lethal action would only be allowed as long as it is conducted in support of the mission (following the rules of IHL and LOAC). Basically, even when a robot is no longer under human control, a human would then still in the loop, as commanders would have defined the mission for the autonomous agent. Just like soldiers must abide by the laws of war, so must the robot abide by its programming to follow the same. The soldier is prescribed to follow IHL and laws of war through training, the robot is instead prescribed through programming and coding (Arkin 2017, 35-36).

Another prominent and seemingly representable understanding of autonomy, which implements perceptions from several other viewpoints in the field of autonomous technology, is delivered by Poul Scharre (2015). Scharre describes autonomy in its plainest shape as the ability of a machine or software to perform *any* task without human input. The machine needs only for a human actor to press its "start button", and then it will simply function and perform its designed task on its own (Scharre 2015, 8). Scharre argues that this is why some experts will talk of a Roomba[10] on the same level as science fiction's Terminator[11]. The perception is then, that these two robots only differ on their level or degree of autonomy. And this is where Scharre surprisingly also criticizes this interpretation of autonomy:

[10] An "*autonomous"* vacuum cleaner, common to some households.

[11] Terminator is a sophisticated autonomous killer robot from the science fiction movie of the same name: 'Terminator' from 1984.

> *"Writers or presenters on this topic often articulate 'levels of autonomy', but their levels rarely agree, leading a recent US Defense Science Board report on autonomy to recommend rejecting the concept of 'levels' of autonomy altogether."* (Scharre 2015, 8)

But Scharre does not abandon this line of thought completely, by gathering insights from other experts of the field and complementing these with his understandings, Scharre instead readjusts the understanding of autonomy away from levels and degrees or other definitions and instead thinks of autonomy in three distinct but concurrent dimensions or axes (Scharre 2015, 9). These dimensions are all independent of one another and autonomy should not just be measured or understood using one dimension, but rather using all three co-dependent yet separate dimensions. Scharre's argument is then, that the discussion on autonomous systems actually follows three completely different concepts which are being discussed as one and the same, and only by separating these can an actual foundation for the discussion be built (Scharre 2015, 9; Nørgaard 2019b, 99). The first dimension is the (1)*the human–machine command-and-control relationship*, which defines the relationship between human and machine. In this dimension we find machines that can perform actions or functions for some time, before stopping, having to receive new human input. As such, they are often referred as '*semi-autonomous*', or with the term '*human-in-the-loop*'. Other autonomous machines that can perform entirely independent, but which are constantly monitored by a human actor, who can intervene or terminate the machine's task at any given time are instead referred as '*human-supervised-autonomy*' or '*human-on-the-loop*'. Finally, autonomous machines which can perform tasks entirely free of human control or supervision, and which humans are unable to stop or intervene are then referred as '*fully autonomous*' or '*human-out-of-the-loop*' (Scharre 2015, 10). With these terms, the measure of human control over the autonomous system or machine is also defined and related to what is called the SDA-loop or *Sense, Decide,* A*ct*-loop. This loop describes the process which comes before a system can conduct a function or perform a task (Nørgaard 2019b, 100).

The second dimension is named (2)*the sophistication of the machine's decision making* and revolves the machine's or system's own decision-making (Scharre 2015, 10). This dimension takes a more technical approach to autonomy, as it relates to a system's ability to exercise "*control over its own behavior (self-governance) and deal with uncertainties in its operating environment"* (Boulanin 2016, 3). The autonomous technologies are then sorted into three or four categories or degrees of autonomy (depending on the author) varying on their functionality and ability to operate independently from human operators. The categories can be varied, from 'Remote-controlled' and/or to 'automatic' to 'automated' to 'autonomous' and/or to 'intelligent' (Scharre 2015 10-11; Boulanin 2016, 3-4; Nørgaard 2019b, 102-103). But, these degrees or boundaries should not be understood as clear or static, their complexity and interconnectedness may lead to different understandings and interpretations of each category, and different experts may very well disagree on what any category entails (Scharre 2015, 11; Boulanin 2016, 4).

The third and final dimension is then called (3)*the type of decision or function being automated.* This dimension relates in many ways directly to the discussion of AWS. It focuses on (the) specific function(s), which are automatic, automated, or autonomous in an autonomous system (Nørgaard 2019b, 105). This dimension stresses the important element of investigating which function(s) in the machine or system is autonomous or likewise, as different decisions are profoundly distinct in their complexity and risk.

> *"A mine and a toaster have radically different levels of risk, even though both have humans 'out of the loop' once activated and both use very simple mechanical switches. The task being automated, however, is much different. Any given machine might have humans in complete control of some tasks and might autonomously perform others. For example, an 'autonomous car' drives from point A to point B by itself, but a person still chooses the final destination. The car is only autonomous with respect to some functions. From this perspective, the question of when we will get to 'full autonomy' is meaningless. There is not a single spectrum along which autonomy moves. A better framework would be to ask which tasks are done by a person, and which by a machine, and what is the relationship of the person to the machine for various tasks. Thus, it is more fruitful to think about 'autonomous functions' of systems, rather than characterizing an entire vehicle or system as 'autonomous'."* (Scharre 2015, 11)

Essentially, this third dimension's more 'functional' approach to autonomy does not make it incompatible with the two earlier dimensions, rather it distinguishes and argues that referring to autonomy as a general attribute of a system is inexact and perhaps even redundant. It is the tasks or functions that are being completed autonomously by a machine or system, which are important. As such, the discussion of degree or level of autonomy inherent in the machine or system becomes superfluous. Autonomy is then best explored relative to the functions being executed. This is also why some autonomous functions in weapons systems can be developed and utilized without presenting greater ethical, legal, or strategic risks (such as: navigation, information reduction, assisted steering, etc.), while others instead cause concern and restraint (such as: targeting, attacking, complete control of own movement, and so on) (Boulanin 2016, 4).

But what can we use this information for, and why is it important to understand all of these different discussions and arguments across the field of autonomous technologies? It is extremely important to understand the negotiations among relevant social and scientific groups to fully understand why and how technologies might be or are being used. Technologies that are implemented does not have their actual usage predetermined by nature or any other factors. It is impossible to predict in advance how users will utilize any technology. During development, design, and implementation of new technology, many factors may change in response of outside factors including, but not limited to, changes in funding, historical events, changes in environment or social and/cultural changes and much more. The technologies that succeed in being adopted and utilized are the outcome of complex negotiations and translations among many different actors and/or actants, including engineers, scientists, politicians, manufacturers, public and more (Noorman & Johnson 2014, 51). There are many different programmes being developed which attempt to problematize the discussion in a specific manner. By following these attempts, and their differing definitions of autonomy across the field, an entryway can be created into the network(s) of actors and actants alike surrounding autonomous weapons systems. By investigating how these different actants relate to autonomous weapons systems and what *they* relate to AWS, can an understanding of the actual world of AWS be uncovered. The Actor-Network perspective allows for the realization that the relations between actants are not static, but are instead constantly reforming and changing. The technology becomes an active part of the discussion and assumes both roles of being a moral agent, creating stability in the chaos of conflict, while also risking developing into a herald of anarchy and uncertainty. Both agents or representations are very real in their respected networks and should be understood as such. This is the empirical constructions which Actor-Network Theory's lens allow us to perceive. By outlining and mapping these empirical networks, differing attempts of stabilization and or destabilization can be revealed and/or negotiated.

5.1 The Value of Definitions of Autonomy - and the Concepts that Follow

So far I have been focusing merely on some of the more common experts' definitions of autonomy, but as the sharp reader might have noticed, some additional concepts emerged during this outlining of autonomy, some of which I shall return to shortly. But, while it is still not my goal to select or decide on which of these definitions or terms that might best define autonomy, my goal is instead to point out how these different definitions open the discussion or "worlds" surrounding AWS in various ways, and also leads to different limitations but also potentials. Depending on how autonomy is understood and perceived, interpretations emerge and are utilized in numerous ways. These different utilizations might be political, social, moral, and so on. As perceptions and interpretations vary, so does the different landscapes or "worlds" surrounding autonomous weapons systems adjust. It is not only experts who vary on their understandings of autonomous technologies, rather, even countries, big organisations and more use these varying perceptions on autonomy to establish their own definitions, which have great impact on technopolitical effects around the globe (Nørgaard 2019b, 87-113). As such, I feel it is relevant to also include some of the more policy-oriented definitions and understandings. The most used description of autonomy is provided by the US Department of Defense (USDOD) which states that an autonomous weapons system is (Nørgaard 2019b, 106; Meier 2016):

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

This definition from the USDOD is quite broad in the spectrum of definitions. It would easily include anything from missile-systems to torpedoes to mines to fully autonomous killer robots because of its lack of specifications or limitation on what it entails. It is likely because of this, that many opponents of AWS also devote themselves to this definition in their attempt for a complete ban on autonomous weapons systems (Nørgaard 2019b, 106). If we instead focus on NATO Multinational Capability Development Campaign(2014) (hereinafter referred as MCDC), the definition of autonomy is instead limited somewhat with the addition of "true autonomy", which is ascribed to humans and other intelligent life, and "autonomous-like" functions, which is what we might find in machines. MCDC instead takes a more functional approach to autonomy in line with Scharre's (2015) third dimension, that focuses on 'the type of decision or function being automated', and states that:

> *"True autonomy should be considered as an intrinsic property of sentient and intelligent creatures. Machines therefore are not autonomous in a literal sense, but may exhibit 'autonomous-like' functions, relative to a particular level of human control and situational context… Second, given that a machine or system in totality cannot be autonomous in a literal sense, the term 'autonomous system' is used with the unstated assumption that not all parts or functions of the system exhibit autonomous-like behavior."* (As cited in Nørgaard 2019b, 106: MCDC 2014, 8)

As such, MCDC focuses rather on the specific function of a weapons system than on the characterization of the system as fully autonomous by itself. The International Committee of the Red Cross has adopted a similar functional interpretation of autonomous weapons systems, where they focus on the critical function(s) instead of the whole weapons system:

> *"The ICRC has defined autonomous weapon systems as: "Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention."* (ICRC 2016, 1)

The ICRC views this classification as a slightly broad definition, that allows for some existing weapons systems to be included while also enabling consideration of other emerging technologies, either in development or in use which calls for concerns under IHL (ICRC 2016, 1-2). Essentially, ICRC has abandoned any distinction between autonomous and automated. The focus, they argue, should rather be on the decisive critical functions, such as the functions that can independently conduct decision-making on life and death (Geiss 2015, 6-7). As such, it can be argued, that the technological definitions are generally cast aside in favour of a more open but focused understanding, which can then focus on the direct human-technology relationship. What becomes pivotal is then instead the ethical and legal standpoint on which decisions that might be delegated to machines and which decisions must stay within the hands of human operators (ibid.).

Another broad definition of autonomy is stated by the Danish Ministry of Defence, in the book "Forsvarets Militærmanual"[12]. The definition offered in this manual seemingly includes both automated and autonomous functions as similar capacities and includes functions such as: information gathering, receiving- and converting data for GPS- and/or shooting range properties, self-governing firing functions, moving and changing of positions, and so on.[13] The Danish military's definition builds itself on top of the potentials for furthered precision and effectivity, assumed to be inherent with autonomous systems. The Danish Defence's definition presents autonomous systems as something which seems *safe* and most importantly: *predictable*. Following this understanding, many autonomous weapons systems are already in use in accordance with the laws of war, and it would be easy to negate argumentation against using autonomous technologies (Nørgaard 2019b, 107-108).

On the other hand, the United Kingdom makes a clear distinction between automated- and autonomous systems. An automated system is one that respond to inputs through the use of sensors, and which is programmed to logically and predictably follow a predefined set of rules in order to arrive at one or multiple intended outcome(s) (Bode and Huelss 2018, 399). In contrast, the autonomous system is then defined as being:

---

[12] English: the Defence's Military Manual

[13] *"Ved autonome (autonomous) systemer forstås, at det enkelte våben er i stand til at positions- og retningsbestemme sig selv ved hjælp af elektronisk (ofte GPS - baseret) udstyr. Autonome systemer kan desuden beregne egne skuddata ud fra digitalt modtagede målkoordinater – eventuelt direkte fra observatøren – og omsætte disse til skuddata. Autonome systemer er normalt selvkørende systemer, men der findes desuden trukne systemer med indbygget retnings- og stedbestemmelsesudstyr samt digitalt ildkontrolsystem. Autonome systemers fordele består i deres øgede hastighed ved stillingsindrykning og klar-til-skud, deres øgede hastighed og lavere fejlmargen ved beregning af skuddata samt mindre mandskabsbehov."* - Knudsen 2016, 578.

> *"Capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be."* (As cited in Bode & Huelss 2018, 399: UK Ministry of Defence 2017, 72)

With this definition, the UK instead creates a clear distinction between automated and autonomous, based on autonomous systems being able to independently understand higher-level intent, and being able to appropriately make decisions leading to desired outcomes. This definition leaves many if not all existing weapons systems within the spectrum of automation and creates rather difficult requirements for creating actual autonomous weapons systems. By making autonomous systems something which is highly unattainable by today's standards, the UK definition essentially eliminates many of the challenges issued by organisations such as the ICRC and instead highlights the existing weaponry as all belonging within the spectrum of automated (Bode & Huelss 2018, 399).

It is interesting to note, how the definitions differ in consequential aspects and in what manner e.g.: the UK definition diverges from the definitions given by NATO MCDC which differs from the USDOD which differs from the Danish Defence Ministry and so on. But, even though the different actors have all adopted their own interpretation of autonomy, which leads to different concepts, ethical concerns, and legal justifications, this does not seem to slow down their reciprocal talks and negotiations on common criteria and definitions for autonomy (Nørgaard 2019b, 109). In fact, even though the talk is based around the definition of autonomy, the focus seems quite often to be on the concepts surrounding it instead.

5.2 Meaningful Human Control – A Moral Inscription Device and A Translation Process

But how then are the negotiations continuing and even gaining pace, when none of the actors speak of the same definitions and interpretations? This may be due to fundamental emerging metaphors and concepts, which can help unite differing policies under one or a few proposition(s). Especially the concept of Meaningful Human Control (MHC), is something that emerges in many different writings across the field, but also (and often in relation) conceptions such as human-in/on/out of-the-loop and Sense-Decide-Act-loop, as well as some universal notions on decision-making. These concepts act as intermediaries or more accurate as "inscription devices" of the discussion. They are something, which the different actors can mostly agree upon, even if they do not agree on a universal definition of autonomy (Horowitz & Scharre 2015; Roff & Moyes 2016).

MHC has gained a role as a dominating concept in the negotiations surrounding autonomous weapons systems. The notion of MHC has been developed as a result of two unacceptable premises within the discussions at the United Nations' Convention on Certain Conventional Weapons (UN CCW)[14]: Firstly, that "*a machine applying force and operating without any human control whatsoever is broadly considered unacceptable."* And secondly, that "*a human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense."* (Roff & Moyes 2016, 1). The fundamental idea, that autonomous weapons systems should be handled with meaningful human control seems understandably appealing to most parties. In many ways the concept's semantical structure makes it uncomplicated to apply. As, who would argue against the idea that a weapon is put under "meaningful" and/or under "human control?". The concept has attained a function as a moral inscription device used in discussions for control over autonomous weapons systems (Nørgaard 2019b, 110-111). But in its success, some problematic avenues has arisen, as no one has managed to figure out what constitutes 'meaningful human control', leading to new questions such as, 'what does "meaningful" entail, and how is it even recognized?'. Almost ironically, discussions on

[14] UN CCW is a convention which "*seeks to protect civilians from the effects of weapons used in an armed conflict and to protect combatants from suffering in excess of that which is necessary to achieve a legitimate military objective*." https://www.icrc.org/en/document/1980-convention-certain-conventional-weapons#.VKkpP2SG-rY
See: https://www.un.org/disarmament/publications/more/ccw/ for more.

a definition of meaningful human control has initiated as a result. Intrinsically, as Horowitz and Scharre argues (2015), this concept is a key element to creating some consensus on the debate of AWS, but it risks acting as an empty cliché, bringing the discussion into yet another philosophical loop or worse (Horowitz & Scharre 2015, 6).

The notion of MHC is in many aspects a concept which actors can unite behind because of its inherent moral performativity. It restructures the discussion on AWS in such a manner, that the arguments of "free-roaming killer robots" are removed or contested to the point of irrelevance, - as with the appliance of MHC such robots would fundamentally never be able to exist, as AWS would be subject to meaningful *human* control. MHC's broadness also allows for the inclusion of both existing "*borderline*" autonomous and future autonomous weapons systems, while also being abstract enough to allow actors to develop separate yet interconnected interpretations across countries and organisations. Of course, this way of constructing the discussion also leaves the functions of AWS in the hands of politicians, as they will have to unitedly define *how* proper meaningful human control would be enacted, as well as which actions must be applied to accomplish this in a proper fashion. The concept risks becoming a political practical tool for enabling development of autonomous weapons systems, but which in essence does not represent either a unified limited usage of AWS nor a universal interpretation to use in accordance with international conventions or polices (Horowitz & Scharre 2015, 6-14; Roff & Moyes 2016; Nørgaard 2019b, 111).

For many of the same reasons, several experts are now attempting to create some formulations or universal classifications for use, when applying MHC to an AWS. Roff and Moyes (2016) formulated several key elements, that must be accounted for, when discussing meaningful human control over an autonomous weapons system in a socio-technical setting. The technology must be '*predictable, reliable, and transparent'*, while providing '*Accurate information'* for the user, both in terms of operation and function. It must also allow for *timely human action* and have embedded a potential for '*timely intervention'* in case of emergencies. Lastly, the expected use of the technology must align with the actual use of the technology, so that '*accountability'* can be established in a proportional manner. In other words, the technology should act as expected, - which could be argued to relate to the first criteria of predictability (Roff & Moyes 2016, 1-3). Finally, Roff and Moyes also stress that the context in which the technology is being utilized must also be accounted for. Here they divide the context into three separate yet interconnected layers: Ante bellum (before conflict), in bello (in conflict), and post bellum (after conflict) (Roff & Moyes 2016, 3-6). Though they do not view their list as exhaustive, they do argue that this way of addressing MHC is a must for creating a unified use of the concept.

Horowitz and Scharre have instead formulated three universal criteria for establishing meaningful human control: *Accountability, moral responsibility,* and *controllability.* Common to these three criteria is that they build on an understanding that technology works in collaboration with humans, as opposed to being an opposition to each other (Nørgaard 2019b, 111). These criteria are quite similar to Roff and Moyes' but are broader and less specified. But all these underlying criteria for fulfilling the meaningful human control over an autonomous weapons system can be viewed as inscription devices used to further specific *action* and hermeneutic interpretations. Understood in a more Actor-Network manner, they can be viewed as part of a translation process that attempts to form the discussion on the usage of AWS.

MHC can then be understood as a prime example of a translation process being initiated by several actants. The UN CCW established this notion to overcome the *problematization* which they framed and believed to be two unacceptable premises when utilizing AWS[15]. In many ways, MHC can then be understood as an attempt at creating *interessement*, locking the discussion and various actants in place. By stating that meaningful human control had to be applied, they managed to introduce this new actant into the various networks surrounding AWS, but rather than creating stability and enforcing its role in the network, the notion was itself challenged. By introducing the notion of MHC to the discussion of AWS, the contested network did not stabilize, rather new negotiations and actants have emerged in the form of the above-mentioned criteria, which becomes part of an underlying network surrounding MHC. It could then be argued that experts, organisations, and state-actors who start to discuss the notion of MHC have become enrolled in the network with MHC. But how these actants relate to each other has yet to be tested and fully negotiated, and as such, interessement has not yet been fulfilled. But this relates well to how

[15] That a *"machine applying force and operating without any human control whatsoever is broadly considered unacceptable."* And secondly, that "*a human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense."* (Roff & moyes 2016)

Callon describes the process of translation as enrolment is dependent on interessment, but interessment is also only a success if enrolment is achieved.

Understanding and following how moral conceptions, such as meaningful human control emerges and are being performed across networks, becomes principal in understanding the negotiations surrounding autonomous Weapons Systems. It is also apparent that the black box containing autonomy has been opened and disassembled to such a degree, that its fundamental usage is near worthless. It has become a term of nothing and everything, and as such, it is only by describing, formalizing, and relating the term of autonomy in new ways, by adding conceptions and metaphors, that its value can be reestablished. As such, behind the term of autonomy is a vast network of various degrees of stability and instability, and it is constantly and justifyingly being challenged and destabilized because of its inherent potential should it ever become stabilized as a finite actant across networks.

The perspectives presented in the above sections does not just display hypothetical worlds, but also empirical constructed worlds. The presented representations reflect the view of experts, organizations and military practitioners who perform and express these definitions, concepts, and "worlds". Many of these actants relate directly to each other either in collaboration, augmentation or opposition. The automated machines or systems are attributed morality in the same manner as humans. This makes it possible to perceive both humans and non-humans on the same level and address these actants in the same manner. However, MHC has become an important actant when ascribing morality to autonomous technology, while still contested, this concept seems to have successfully started a translation process of interessement and enrollment, though it is yet to be seen exactly how its role will be defined.

5.2.1 A Technological Agent of Change

As mentioned in section 5.2.1, the technology has been assigned the role of a moral agent, with the inherent ability to either stabilize or destabilize the very foundation of ethical warfare. Whether advocates or opponents of AWS, this moral agency is ascribed similarly to AWS, which is performed through its relation to the described context(s). This only restates the importance of exploring the contexts which the technology is being presented as part of. Depending on the context; ethical, legal, social, political, and other implications change. What is legitimate and what is not changes depending on the construction and relations of the network. This is an important realization to properly understand and analyse human and non-human agency. Both human and technology have numerous inherent networks which overlaps when they encounter each other, and which evolve into separate networks as human and technology combine through performing these relations.

Currently, the Actor-Network lens have helped reveal insights on *how, where,* and *why* the debate on autonomous weapons systems is performed. I have currently shown that autonomous weapons systems are being performed as moral agents across various interconnected networks. Because of autonomy's inherently devalued definition, new concepts have been introduced to the network in attempts to stabilize relations and framings by means of translation processes. Here meaningful human control is currently the most dominating example of an attempt at problematizing and framing the discussion on autonomy. But because of MHC's broad interpretations, both opponents and advocates have embraced the concept and MHC are in risk of itself being repurposed in unintentional ways. For this reason, categorizations have been developed to stabilize MHC and define what is meant by meaningful human control. Furthermore, the context in which these discussions are being presented are revealed to be essential, as they can lead to different narratives on what is legitimate and what is not legitimate. The polices, that have been presented above, all represent different actants part of the networks involving autonomous weapons systems. Human operators and pilots who utilize autonomous weapons systems also become part of these networks. As such, these operators and more are not neutral when they actualize their usage of this technology but are already part of a translated network filled with moral inscriptions. Technopolitical realities are therefore an essential part of the relation between pilot/operator and drone/robot or other.

According to Nielsen and Bollmann (2019) an example of the mutual constitution of technology and policy can be perceived in the usage of the Remotely Piloted Aircraft (RPA): Global Hawk[16]. In an interview with a chief military officer in NAGSF[17] they discovered how the usage of the RPA was associated and framed

[16] Global Hawk is an unmanned reconnaissance airplane most famous for its usage within the NATO alliance for surveillance and information gathering. See e.g.: https://www.nato.int/cps/en/natolive/topics_48892.htm for more.
[17] NATO Alliance Ground Surveillance Force

in specific ways (Nielsen & Bollmann 2019, 123). The technology was described as a "strategic platform", while similarly being able to "send political messages", but it also affected the decision-making processes all way from the operational to the tactical level (Nielsen & Bollmann 2019, 123). While Global Hawk is not autonomous per se, it has, like the Predator, Reaper, and other remote piloted aircraft certain simple autonomous capabilities. As such, they are the current closest comparative to how future autonomous weapons systems will come to used and perceived. Another of Nielsen and Bollmann's highlighted interviews involve an operator of a Global Hawk who describes their relation to this drone. The operator clarifies the importance of recognizing the Global Hawk as a Remotely Piloted Aircraft (RPA) as opposed to an Unmanned Aerial Vehicle (UAV). The reason being, that the operator believes that otherwise, there would be an underlying assumption that the drone flew on its own, - being autonomous. The operator attempts to distance the drone and their own operation from this assumption, - that the Global Hawk is autonomous. Rather describing it as a "*regular plane, where the cockpit has just been placed on the ground*" (Nielsen & Bollmann 2019, 131-132). Though the operator admits that the "plane" does have autonomous capabilities in case of disrupted communications. The operator views this more of an ethics-dilemma about predictability, which the operator somewhat swiftly disqualifies expressing: "*Even if there were pilots physically in the plane* [Global Hawk], *we are not even at that point*." (Nielsen & Bollmann 2019, 132).

The construction of the Global Hawk as a remotely controlled airplane as opposed to an autonomous system shows the importance of the revolving constituted understandings, interpretations, and discussions. It returns the discussion to the boundary between automation and autonomy. The definitions from which we define autonomy becomes increasingly important in such situations, as it changes our very view and perception on a technology such as the Global Hawk. Currently the technology is subject to several moral, legal, and political polices, but these would change quickly if the Global Hawk were to be suddenly viewed as autonomous. This would reshape the whole network and destabilize many current actants and initiate a whole new reconstruction of both human and non-human actants. As such, the operator defends the currently stable network in which they have been enrolled. RPA becomes and active moral inscription, which, in a similar manner as MHC, is performed as justification for using the technology.

6. Moralizing *through* Technology

As we come to realize the performativity of these moral inscriptions to justify or oppose autonomous weapons systems in an actor-network setting, and how the context becomes crucial to understanding different actors' enrollment in different programmes, several fundamental moral and social controversies emerge. I have outlined some underlying controversies as well as their interconnected relations, and I have discovered, that one of the major controversies, that have emerged during this investigation, revolves around responsibility and accountability, as in - 'how can responsibility for specific actions be delegated to humans in a socio-technical system containing autonomous technological entities?'. And even though ANT have helped describe this controversy, and especially how the differing definitions of autonomy changes the view on moral responsibility, it has not provided any evident solution or method to overcome this controversy altogether. I therefore substitute the material-semiotic lens in favour of the embodied lens from postphenomenology. By building on top of the ANT-analysis and the relational understandings, it is my belief, that the postphenomenological-lens will be able to provide another dimension, which is able to go closer into the embodied and mediated actions of humans employing AWS. By doing this, I wish to provide this analysis with a two-dimensional perception into the interconnected worlds of autonomous weapons systems.

As I have shown in my investigation, autonomy is not a *specific* concept. And even if I adopted any of the represented definitions, it becomes hard to precise where and how to appoint responsibility and/or accountability. Noorman and Johnson (2014) demonstrates how responsibility for these technologies have to do with decisions, strategies, employment, design, et cetera, and how practical questions are likely to follow if something were to go wrong: Was the design adequate for the task? Were there any inherent flaws? Were the systems tested properly? Did those who utilized the systems make mistakes? Was the technology used correctly? Was the technology used in illegal ways on purpose? Such questions will likely be asked in case of failures or unintended use (Noorman & Johnson 2014, 60). These questions will demand many different perspectives from different fields to fully answer and understand, as such I do not seek to answer all of the questions above, instead I embarked on this analysis with the goal of creating a framework, which can help pragmatically explain the challenge of users' moral responsibility in relation to autonomous weapons systems, and how this phenomenon is neither entirely human nor wholly non-human.

Depending on the definition of autonomy many current developing weapons technologies could be included as autonomous weapons systems, and though many of these are still not operational or only prototypes, their research and development are constantly proceeding. AWS is also not just a single platform, but can take many different forms, from singular to swarm, from land; to air; to sea, - many platforms, existing or developing, are debatably only stages away from attaining increasing functional autonomous properties (Bode & Huelss 2018, 400). The US Office of Naval research are testing and have successfully operated several autonomous unmanned "swarming boats" during exercises. Likewise, the USDOD have announced successful testing of a "micro-drone swarm" in air, expressing that the drones demonstrated collective decision-making, adaptive formation flying, and self-healing (Bode & Huelss 2018, 400-401). But as I have already presented much earlier, these developments are not confined to the US. Examples such as the operational South Korean Samsung SGR-A1 autonomous sentry robot or the Israelian Harpy system, a sensor-fused loitering munition, provide a clear picture that these technologies are already in use around the globe (Boulanin 2016, 8; Bode and Huelss 2018, 402). Whether these technologies are autonomous is a matter of negotiation and something which I will relate to my earlier analysis above. But, if we agree that these technologies have autonomous capabilities, these capabilities are still regulated by human operators involved in the procedures, usage, and decision-making. It is still a human action that activates these technologies, and for most of these AWS humans either have full control of whether the system engages, or supervises the technology with the intention of intervening when necessary (Boulanin 2016, 7-9; Bode and Huelss 2018, 397-402).

But whether we are discussing the reconnaissance aircraft 'Global Hawk' or the autonomous loitering munitions-system 'Harpy', these technologies are more than just neutral instruments. They actively co-shape the world in which they exist. They essentially change how humans perceive and act, and as users do not just perceive the world, they perceive the world *through* the technology (Verbeek 2011, 4). Though this makes it important to note the use of the term "co-shape", as otherwise the discussion risks pertaining a technological deterministic perception, where the technology becomes dominating and determining over human action, and as a result also reducing human responsibility and moral accountability. Instead, when users either supervise or use weapons technologies, their perception of reality and the information that they base decisions on are created through a constituted reality in combination with the technology. An example is provided by Timothy Cullen (2011) who conducted an empirical study on the usage of the

MQ-9 Reaper. While the Reaper is, as commented on in section 1, not autonomous per se, it does have some autonomous features and many view it as a precursor to future advanced autonomous weapons systems. The MQ-9 Reaper is an armed multipurpose, medium altitude, and long endurance remotely piloted aircraft. It is employed both as a weapon and as an intelligence collection asset (US Air Force, 2015). During Cullen's interviews he was told by pilots under training, that they felt like they were being taught how to "fly the matrix". The torrent of information, symbols and dialogue going across the interfaces of the screens were the pilots' only access to the world as provided by the sensors of the MQ-9 Reaper (Cullen 2011, 117).

> *"For the untrained eye, it was enough to make a person sick to their stomach. Pilots learning how to fly the aircraft said it felt like they were "flying the matrix." They often got lost in a cacophony of cascading symbols and overlapping conversations. Eventually, Reaper operators found their way, however, and their view of the system evolved. Experienced Reaper crews learned how to configure and coordinate the matrix of verbal, textual, graphical, and geographical representations in the ground control station and how to use them as tools to identify threats and eliminate them*." (Cullen 2011, 117).

This is a core example of the way in which the world is perceived through technology. But the pilot not only perceives the "real" world through the screens of the Reaper, the pilot performs actions, makes decisions, and is connected to the world through the technology. Their moral understandings mediated by whatever the interface might show. This is the mediating role of technology that postphenomenology allows us to investigate. And what becomes especially important in this setting, is exactly this mediation of the users' moral understandings and actions. As Verbeek explains: *"If ethics is about the question 'how we act', and technologies help to answer this question, technologies appear to do ethics, or at least help us to do so*." (Verbeek 2011b, 205). As such, Reaper pilots act upon this co-shaped *morally* mediated understanding when using their tools to "identify threats and eliminate them". Technology becomes a means for humans to functionally *moralize through technology*. Technology is a mediator for our actions and understandings, but it not only enables us to carry out actions that previously were not possible, or hardly possible, it shapes our very experience of things. The MQ-9 Reaper is an excellent example of this. It is through the interface of the screens that pilots are faced with the responsibility of making difficult life and death decisions. They conduct moral decisions on the life of their possible targets, and they even define the world for others through its information-gathering properties, relaying a technological worldview to other actors. At the same time, however, the MQ-9 Reaper does not determine the pilots' actions but is purely informing and assisting as programmed. In the end, it is a constituted perceived reality, which becomes the foundation for moral action. This is why Verbeek views "*moral decision-making as a joint effort of human beings and technological artifacts*" (Verbeek 2011b, 206).

> *"Moral mediation always involves an intricate relation between humans and nonhumans, and the "mediated agency" that results from this relation therefore always has a hybrid rather than a "purely nonhuman" character. When technologies are used, moral decisions are not made autonomously by human beings, nor are persons forced by technologies to make specific decisions. Rather, moral agency is distributed among humans and nonhumans; moral actions and decisions are the products of human-technology relations"* (Verbeek 2011, 53)

When perceived in this manner, it then becomes clear that moral responsibility cannot be solely technological, and that humans *are* the bearers of responsibility, but neither is it entirely human, instead, it is a combination. It is a result of a convoluted mixture of humans actualizing their morality through technological artefacts. Yet the discussion could then easily be derailed by only focusing on design. As it could be argued that it is the interfaces, shapes, languages, etc. of technology which mediates human perception - and thereby also action. Depending on how technology is then designed, the mediation changes. This conclusion would lead to tremendous ramifications for engineers and designers, as they would then be solely responsible for the technological artefacts and the underlying meditation that they have created. But this is only a half-truth. The notion of mediation illustrates, that designers should be aware of the inherently moral activity of *moralizing* through *technological design* (Verbeek 2011b, 207; See also Verbeek 2011). But technology can be and *always will be* used in a variety of ways. Cullen describes during his observations on the usage of the MQ-9 Reaper, operators changing their interfaces and repurposing how the autonomous functions are being used. The pilots and sensor operators learned how to "overcome the operational assumptions" of the engineers (Cullen 2011 118).

"For Reaper pilots and sensor operators to employ the system with respect, they had to learn how to coordinate and distribute data among a hodgepodge of display and input devices and how to overcome the operational assumptions engineers used to build the system. General Atomics engineers initially designed the aircraft to fly autonomously for the bulk of a mission, but pilots modified the ground control station and their procedures to share aircraft control with the autopilot in order to maneuver more quickly and destroy a target at a specified time. Raytheon engineers designed the aircraft's sensor ball to adjust the iris of its cameras automatically, but sensor operators developed procedures and a frame of reference to adjust the cameras' settings manually in order to investigate obscured areas of a scene and to create images that seemed more natural. Reaper pilots and sensor operators were able to employ the weapon successfully despite the unanticipated practices of the operators and the seemingly haphazard construction of the ground control station because they acted as the "malleable and adjustable coordinating tissue" of the system." (Cullen 2011, 118-119).

This observation from Cullen illustrates the inherent multistability in technology. The engineers who designed the Reaper's different functionalities most likely designed it with specific actions in mind. But the pilots and operators instead readjust the technology to their preferences and demands. These unintended or unexpected adjustments of the technology might completely change the expected mediation of the technology, and as such, designers cannot be certain that their inventions and creations are utilized as originally anticipated. Designers can therefore not be viewed as fully responsible for the usage of their creations, and this only becomes even clearer in scenarios where the 'advanced' autonomous technology works in combination with human operators. It would be an incalculable task to completely expect potential unintended use. As because of the very self-governing nature of future autonomous technology, it is only natural to assume, that their possible usage will be exceptionally multifarious.

What we then come to realise, is that when operators use the MQ-9 Reaper, they do so with preunderstandings and intent. As my ANT-analysis showed, the operator and/or pilot is not neutral, but is an enrolled actor in the network surrounding AWS. As such, they carry a programme translated within the network with them, which also affects their usage and relation to the technology. As I showed with the example from the operator of a Global Hawk from Nielsen and Bollmann (see section 5.2.1), the constituted world of the network changes how we perceive moral controversies. The moral inscriptions translate moral programmes that change how humans understand their actions. The different autonomous weapons systems become a gateway into putting these moral perceptions into action. As such, the autonomous technology *is* a moral agent, and together with the contextual understanding and moral inscriptions used to translate morality, pilots and operators *perform* this morality through the technology. But the technology also comes with an inherent moral design, which only comes into being, through usage in combination with a human operator. This is what Verbeek and Ihde calls technological intentionality (see section 3.2 and 3.2.1). By acting through the technology, a hybrid intentionality emerges.

7. Discussion

7.1 Future Research and Possible Avenues

Throughout my study, I have gained several exciting insights into the field of autonomous weapons systems. I have also come across some gaps, that I believe will help further the discussions on AWS in a fruitful manner. I will firstly discuss *how the different representations limits and opens the field* in various ways, secondly I will *discuss combining Actor-network Theory with moral understandings from postphenomenology* as a means to further investigate this field, and lastly I will conduct an *inspired discussion on the innovation of technology in general*.

7.1.1 The Different Representations' Potentials and Limits

As I presented in section 5.1 and 5.2, how autonomy is understood, changes the very essentials of the discussion. My analysis seems to suggest that the discourse about autonomy and moral responsibility is an interconnected negotiation. The field surrounding autonomous weapons systems is filled with different translations, each attempt at enrolling the other actants into its programme, - either as allies but also as opponents. And in every translation I found that several inscription devices emerged. Meaningful Human Control and even the conceptual difference between unmanned aerial vehicles and remotely piloted aircraft are used as moral inscription devices to further different programmes and stabilize other actants.

My analysis also showed how morality was performed across networks, and how AWS was enrolled as a moral agent. Depending on the translation process, the technology is either attempted to be locked as an agent of unity and ethical pureness or as an agent of chaos and destabilization. The important understanding is that the technology performs moral understanding and also translates further programmes itself. It is given the status of a moral entity on the same level as humans. As such, pilots and operators are also involved in its performed morality. In section 5.2.1 I presented the example of a Global Hawk operator who distanced themselves from the concept of unmanned Aerial vehicle, simply because of the underlying connection between UAV and autonomy. Instead the operator allied with the concept of RPA, as this concept instead suggested that the aircraft was under full human control.

This goes to show, that there is much more to learn from empirically interviewing and observing interaction with autonomous technologies. Where this study has revealed much of the field's understandings and has mapped experts', states' and organsiations' interpretations of autonomy through the usage of literature, I argue that there would be much to learn from studying the technologies in person and in usage. Hidden understandings and, black boxes and tacit knowledge are almost certainly undiscovered and hidden in the networks surrounding AWS, just waiting for an ethnographic discoverer. But, as I have argued earlier in this report, I also believe that actor-network theory can be augmented by adding a simultaneous postphenomenological moral understanding.

7.1.2 Combining Actor-Network Theory with moral understandings from Postphenomenology

During my analysis I attempted to apply a postphenomenological analytical understanding of the moral meditation of autonomous weapons systems. And while I did not follow postphenomenology's human-technology relations, I instead understood the technology's relation through ANT. I discovered that while this does create some obstacles, it also gave my analysis a second dimension to understand the relationship between human and technology. Instead of attempting at choosing which of the theories that were better suited for explaining technological agency, I accepted both in their own right.

When perceiving through an ANT-lens, the technology is understood as being part of underlying moral translation processes and programmes. As such, the technology cannot be understood as a neutral entity, already by being part of a network, it is an agent performing moral agency, which can only be revealed by exploring its relational connection to other actants, human and non-human. Because of the non-neutral nature of such a moral agent, it becomes apparent that this also affects how pilots and/or operators perceive the very technology that they use. But, when applying an embodied understanding and exploring the technology through following the hermeneutic mediated understandings and pragmatic mediated actions, I instead presented how the users come to understand the world *through* the technology. Be it either Reaper or Global Hawk, the technology becomes a gateway to the "real" world, and as such mediating action and perception. Moral action is conducted *through* the technology. This only goes to highlight that the technology is once again not neutral. Instead it has fundamental designed "scripts" or intentionality which co-shapes the world in combination with the user, creating a hybrid intentionality.

On the background of this, my analysis then came to show, that responsibility cannot be solely technological, as it is connected to the user's own morality, which the user actualizes through usage of the technology. But the user's moral understandings are part of the socio-technical network, which surrounds the field. As such, the technology does have agency in the network, as a representation of certain discourses, but it does only gain agency in its embodied form through usage in combination with humans. As such, my analysis has also revealed that the development of autonomous weapons systems does not necessarily imply the conclusion of human responsibility, which some of the more technological deterministic narratives might suggest. Instead I believe that this ANT and postphenomenological combined approach can help further a framework for uncovering how to better distribute responsibility and accountability. I do not view this as a final or completed methodological framework, but it is a start which could act as a foundation for further future research into the distribution of responsibility and accountability in the socio-technical world of autonomous weapons systems.

7.1.3 On the Innovation of Technology

Finally, I wish to bring a perspective on the heterogenous concept of innovation into the discussion on AWS. When technology is being developed, it is often accompanied by the keyword of "innovation". This socio-material concept and its underlying processes are an untold part of the development of drone and autonomous technology. Not much of the literature emphasizes on the design processes or the complicated scenarios which have led to technological development. It seems to disappear in the midst of the many controversies that emerge in the field. But I imagine that some parallels and important experiences can be gathered from examining other prior innovative processes. It is a fact, that describing the process of innovation can be a difficult task. Looking at innovation retrospectively might provide easy answers to why some were successful, and others failed, but what do we do in the case of innovation in the making? - This proves to be a much-complicated scenario (Akrich, Callon, & Latour 2002, 187). Akrich et al. expresses on innovation in the making that:

> *"[I]nnovation in the making reveals a multiplicity of heterogeneous and often confused decisions made by a large number of different and often conflicting groups, decisions which one is unable to decide a priori as to whether they will be crucial or not."* (Akrich, Callon, and Latour 2002, 191)

For this reason, innovation should not be conceived as a deliberate and nor controlled nor linear process controlled by a few industries or innovators. Innovation is a controversial process, where actors clash and negotiate in networks of power, size, and translations. We must therefore not fall into the trap of perceiving innovation processes as purely controlled or guided by logic, order, or rationale. It is a context dependent process, which dwells within situated uncertainties and irrationality as well (Akrich, Callon, and Latour 2002, 196).

Instead of attempting to only understand such processes from the outside, we should strive to understand AWS from the inside as well. Attempting to gain access to the development of autonomous weapons systems and explore the relations, which is inherent in its network. Akrich et al. (2002) argues, that to create successful innovation, one must create interessment and alliances, but that the innovation also depends much on the technical choices, why it is not enough to look either at the social or technícal of an innovation. Instead symmetry in the form of a socio-technical analysis is needed, whether it be an organisational or technological innovation (Akrich, Callon, and Latour 2002, 204). Innovation then becomes a performed in a situated reality, where the social and the technological shape each other (Akrich, Callon, and Latour 2002, 204-205). Following this understanding and perspective, I see a gap in the literature ready to be filled with ethnographic studies of the innovative properties of AWS. This study would not only help further understanding on the negotiation of AWS but would be a perfect case study into the translation processes that follow highly innovative technology.

8. Conclusion

This thesis has explored the socio-technical world of autonomous weapons systems in an attempt at exploring the constituting effects and how this affects the emerging controversies. I conducted this investigation through the usage of a literature study into the field, which then served as my foundation for understanding the many representations of autonomy. The mapping of the definitions of autonomy revealed how the definitions are used in a battlefield of complex negotiation and translation. There are many different programmes being developed which attempt to problematize the discussion in a specific manner. By following these attempts, and their differing definitions of autonomy across the networks. I have investigated how these different actants relate to autonomous weapons systems. The Actor-Network Theory perspective allowed for the realization that the relations between actants are not static, but are instead constantly reforming and changing, and that the autonomous technology are an active part of the discussion and has been ascribed two differing roles as moral agent. One role describes the technology as creating stability and ethical unity, while the other description depicts the technological moral agents as a herald of uncertainty, unpredictability, and militarization. Both agents or representations are very real in their respected networks and should be understood as such. By outlining and mapping these empirical networks, differing attempts of stabilization and or destabilization have been revealed and/or negotiated. My analysis also showed that human operators and pilots who utilize autonomous weapons systems have also become part of these networks. But, the operators are not neutral when they use autonomous technologies but are instead already part of translated networks filled with moral inscriptions. Technopolitical realities are therefore an essential part of the relation between pilot/operator and drone/robot.

Finally, I attempted to apply an Actor-Network Theoretical and postphenomenological combined approach in an attempt to expand the investigation on the relationship between human and technology in this socio-technical world. While I believe that more work needs to be done on this fletching framework, which I have presented in this thesis, I view my combination as a success by itself, which provided me with a two-dimensional perspective into the socio-technical world of autonomous weapons systems. The material-semiotic understanding revealed how the technology shapes our preunderstandings before we even utilize it. It helps highlight the morality which might be inherent in the technology. The embodied understanding delivered an insight into the co-shaped reality of an MQ-9 Reaper, showing how the user and technology combined their intentionality and created a hybrid hereof. It also revealed that the moral responsibility is inherently human, as the technology in action, acts out human morality. Humans moralize through technological means.

9. Bibliography

Akrich, Madeleine 1992. “The de-scription of technological objects.” In ‘Shaping technology/building society’, edited by Bijker and Law. Cambridge: MIT Press.

Akrich, Madeleine, Michel Callon, and Bruno Latour 2002. “The Key to Success in Innovation Part I: The Art of Interessement.” International Journal of Innovation Management Vol. 06. Issue 2 pp: 187–206. https://doi.org/10.1142/S1363919602000562.

Arkin, Ronald C. 2010. “The Case for Ethical Autonomy in Unmanned Systems.” Journal of Military Ethics, Vol. 9, Issue 4, pp: 332-341, December 2010. DOI: 10.1080/15027570.2010.536402

Arkin, Ronald C. 2017. "A robotocist’s perspective on lethal autonomous weapon systems." In ‘UNODA Occasional Papers No. 30, November 2017: Perspectives on Lethal Autonomous Weapon Systems’. United Nations, New York. November 2017 https://doi.org/10.18356/7748aa31-en.

Bieri, Matthias and Marcel Dickow 2014. “Lethal Autonomous Weapons Systems: Future Challenges.” CSS Analyses in Security Policy. No.164, November 2014.

Bode, Ingvild and Hendrik Huelss 2018. ” Autonomous weapons systems and changing norms in international relations.” In Review of International Studies Vol. 44, issue 3: 393-413. July 2018.

Boulanin, Vincent 2015. “Implementing Article 36 weapons reviews in the light of increasing autonomy in weapon systems” in SIPRI: Insights on Peace and Security, Nov 2015/1.

Boulanin, Vincent 2016. “Mapping the development of autonomy in weapons systems: A primer on autonomy”. Stockholm International Peace Research Institute (SIPRI)

Boulanin, Vincent and Maaike Verbruggen 2017. “Article 36 Reviews: Dealing with the challenges posed by emerging technologies”. SIPRI: Stockholm International Peace Research Institute. Dec 2017.

Callon, Michel 1984. “Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay.” The Sociological Review, Vol. 32, Issue 1: 196-233. May 1984. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Coeckelbergh, Mark 2013. “Drones, information technology, and distance: mapping the moral epistemology of remote fighting” Ethics and Information Technology, Vol.15(2,). DOI: 10.1007/s10676-013-9313-6, Mar 2013.

Cullen, Timothy M. 2011. “The MQ-9 Reaper Remotely Piloted Aircraft: Humans and Machines in Action” Massachusetts Institute of Technology, Sep 2011.

Englert, Matthias, Sandra Siebert, and Martin Ziegler 2014. ”Logical Limitations to Machine Ethics with Consequences to Lethal Autonomous Weapons.” Computer Science, Nov. 2014 ArXiv, abs/1411.2842.

Geiss, Robin 2015. “The International-Law Dimension of Autonomous Weapons Systems”. Friedrich Ebert Stiftung. Oct 2015.

Geiss, Robin 2016. “Autonomous Weapons Systems: Risk Management and State Responsibility.” Third CCW meeting of experts on lethal autonomous weapons systems (LAWS) Geneva, 11-15 April 2016

Graae, Andreas Immanuel 2019. “The Cruel Drone: Imagining drone warfare in art, culture, and politics.” University of Southern Denmark. The faculty of Humanities.

Green, Bart N., Claire D. Johnson, and Alan Adams 2006. “Writing narrative literature reviews for peer-reviewed journals: secrets of the trade” Journal of Chiropractic Medicine Vol. 5, Issue 3. pp: 101-117 Autumn 2006. https://doi.org/10.1016/S0899-3467(07)60142-6

Grut, Chantal 2013. “The Challenge of Autonomous Lethal Robotics to International Humanitarian Law.” Journal of Conflict and Security Law, Vol. 18, Issue 1. pp: 5–23, April 2013. https://doi.org/10.1093/jcsl/krt002

Horowitz Michael C. and Paul D. Scharre 2015. “Meaningful Human Control in Weapons Systems: A Primer.” CSNAS Working Papers: Center for a New American Security. March 2015.

ICRC 2011. “International Humanitarian Law and the challenges of contemporary armed conflicts”. Report for the 31st International Conference of the Red Cross and Red Crescent, Geneva, 28 November to 1 December 2011, pp. 39-40: https://www.icrc.org/eng/assets/files/

ICRC 2016. “Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system”. Made for the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) April 2016, Geneva. 11th April 2016.

Ihde, Don 1986. “Consequences of Phenomenology.” State University of New York Print, Jan 1986.

Ihde, Don. 1990 “Technology and the Lifeworld - From Garden to Earth.” Indiana Series in the Philosophy of Technology University Press. Bloomington: Indiana. ISBN13: 9780253205605

Ihde, Don 2015. “Preface: Positioning Postphenomenology”, in Rosenberger & Verbeek’s ‘Postphenomenological Investigations: Essays on Human–Technology Relations’. Postphenomenology and the Philosophy of Technology. Lanham, MD: Lexington Books.

Jensen, Torben E. 2003. “Aktør-Netværksteori – En Sociologi Om Kendsgerninger, Karakker og Kammuslinger.” Papers in Organization 48.

Kastner, Monica, Sharon Straus, and Charlie H. Goldsmith 2007. “Estimating the Horizon of Articles to Decide When to Stop Searching in Systematic Reviews: An Example Using a Systematic Review of RCTs Evaluating Osteoporosis Clinical Decision Support Tools” *AMIA Annual Symposium Proceedings volume*.

2007. pp. 389-393. Oct 2007.Knudsen, Jes R. 2016. “Militærmanual om Folkeret for Danske Væbnede Styrker i Internationale Militære Operationer.” København: Forsvarsministeriet. September 2016.

Kroustrup, Jonas and Finn Olesen 2007. “ANT – Beskrivelsen af heterogene aktør-netværk” in ’Introduktion til STS: Science, Technology, Society’, edited by Casper Bruun Jensen, Peter Lauritzen and Finn Olesen. Hans Reitzels Forlag 1st e-book edition. København 2008.

Latour, Bruno 1992. “Where are the missing masses? The sociology of a few mundane artifacts” in ‘Shaping technology/building society’, edited by Bijker and Law. Cambridge: MIT Press. 1994.

Latour, Bruno and Steve Woolgar. 1979. “Laboratory Life: The Social Construction of Scientific Facts.” Princeton University Press. Sage, Los Angeles.

Latour, Bruno 1990. “Technology is Society Made Durable.” The Sociological Review, vol. 38 issue 1: 103-131. May 1990. https://doi.org/10.1111/j.1467-954X.1990.tb03350.x

Law, John 1992. “Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity.” Systems Practice (5): 379-393 Centre for Science Studies. Lancaster University. August 1992. https://doi.org/10.1007/BF01059830

Law, John 1999. “After Ant: Complexity, Naming and Topology.” The Sociological Review 47: 1–14. https://doi.org/10.1111/j.1467-954x.1999.tb03479.x.

MCDC 2014. “Multinational Capability Development Campaign (MCDC) 2013-2014. Focus Area ‘Role of Autonomous Systems in Gaining Operational Access’”. Policy Guidance: Autonomy In Defence Systems. 29th October 2014.

Meier, Michael W. 2016. “Lethal Autonomous Weapons Systems (Laws): Conducting a Comprehensive Weapons Review.” The Temple International and Comparative Law Journal. Vol.30, Issue 1. Temple University Beasley School of Law. Spring 2016.

Mindell, David 2015. “Our Robots, Ourselves: Robotics and the Myths of Autonomy” Penguin Publishing Group, Viking, ISBN: 9780525426974, Oct 2015.

Nielsen, Thomas Galasz, and Anders Theis Bollmann 2019. “5. Det strategiske perspektiv: Teknopolitiske dagsordner og konfigureringen af det strategiske miljø.” in ’Robotterne Styrer: Militær teknopolitik og risikoledelse i praksis’, edited by Katrine Nørgaard and Søren Sjøgren. Samfundslitteratur 1. edition 2019.

Noorman, Merel, and Deborah G. Johnson 2014. "Negotiating Autonomy And Responsibility In Military Robots". Ethics And Information Technology 16 (1): 51-62. doi:10.1007/s10676-013-9335-0.

Nørgaard, Katrine 2019 "Introduktion: Nye teknologier på en ny kampplads – Mennesker og maskiner i krig" in ʼRobotterne Styrer: Militær teknopolitik og risikoledelse i praksis', edited by Katrine Nørgaard and Søren Sjøgren. Samfundslitteratur 1. edition 2019.

Nørgaard, Katrine 2019b "Genstandsfeltet: En teknopolitisk terrænorientering" in ʼRobotterne Styrer: Militær teknopolitik og risikoledelse i praksis', edited by Katrine Nørgaard and Søren Sjøgren. Samfundslitteratur 1. edition 2019.

Roff, Heather M. and Richard Moyes 2016. "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons." Briefing paper for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016.

Rosenberger, Robert and Peter-Paul Verbeek (Ed.) 2015. "Postphenomenological Investigations: Essays on Human-Technology Relations". Postphenomenology and the Philosophy of Technology. Lanham, MD: Lexington Books.

Sauer, Frank 2016. "Stopping 'Killer Robots': Why Now Is the Time to Ban Autonomous Weapons Systems". Arms Control TODAY Volume 46, Number 8, October 2016.

Scharre, Paul D. 2015 "The Opportunity and Challenge of Autonomous Systems." In 'Autonomous Systems: Issues for Defence Policymakers,' edited by Andrew Williams and Paul Scharre. Capability Engineering and Innovation, HQ SACT, Norfolk United States. ISBN 9789284501939

Scharre, Paul D. 2016. "Autonomous Weapons and Operational Risk." Center for a New American Security. February 2016.

Sharkey, Noel 2009. "Death Strikes from the sky: The Calculus of Proportionality" IEEE Technology and Society Magazine, vol. 28, no. 1, pp. 16-19, Spring 2009, DOI: 10.1109/MTS.2009.931865.

Sparrow, Robert 2007. "Killer Robots." Journal of Applied Philosophy Volume 24, Issue 1. 24.62-77.10.1111/j.1468-5930.2007.00346.x.

Surber, Regina 2018. "Artificial Intelligence: Autonomous Technology(AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats" ICT4Peace Foundation and the Zurich Hub for Ethics and Technology (ZHET). Feb 2018

Thurnher Jeffrey S. 2016. "Means and Methods of the Future: Autonomous Systems." In: Ducheine P., Schmitt M., Osinga F. (eds) Targeting: The Challenges of Modern Warfare. T.M.C. Asser Press, The Hague

UK Ministry of Defence 2017. "Joint Doctrine Publication 0-30.2. Unmanned Aircraft Systems" Joint Doctrine Publication 0-30.2 (JDP 0-30.2). August 2017. available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/640299/20170706_JDP_0-30.2_final_CM_web.pdf

Umbrello, Steven, Phil Torres, and Angelo F. De Bellis 2019. "The future of war: could lethal autonomous weapons make conflict more ethical?". AI & Society Vol. 35, pp. 273–282 Issue: March 2020. https://doi.org/10.1007/s00146-019-00879-x

US Air Force 2015 "MQ-9 Reaper." Air Combat Command, Public Affairs Office. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104470/mq-9-reaper/ - Last visited 09/01/2020

US Department of Defense 2012. "Directive 3000.09. Autonomy in Weapons Systems." United Sates of America, November 21, 2012. Incorporating Change 1, May 8, 2017.

Verbeek, Peter-Paul 2006a. *"*Persuasive Technology and Moral Responsibility Toward an ethical framework for persuasive technologies*"* Paper for Persuasive06, Eindhoven University of Technology.

Verbeek, Peter-Paul 2006b. "Materializing Morality: Design Ethics and Technological Mediation" in Science, Technology, & Human Values 31(3), 361–380. https://doi.org/10.1177/0162243905285847

Verbeek, Peter-Paul 2008. "Cyborg intentionality: Rethinking the phenomenology of human–technology relations." Phenomenology and the Cognitive Sciences vol. 7: 387–395. June 2008. https://doi.org/10.1007/s11097-008-9099-x

Verbeek, Peter-Paul 2011. "Moralizing Technology: understanding and designing the morality of things.*"* Chicago/London: University of Chicago Press. ISBN 978 022 6852911, 978 022 6852904

Verbeek, Peter-Paul 2011b. "7. Designing Morality" In 'Ethics, technology, and engineering : an introduction,' edited by Ibo Van de Poel and Lambér Royakkers. Wiley-Blackwell publishing Ltd., Chichester. ISBN 978-1-4443-9570-9

Williams, Andrew P. 2015. "Defining Autonomy In Systems: Challenges and Solutions." In 'Autonomous Systems: Issues for Defence Policymakers,' edited by Andrew Williams and Paul Scharre. Capability Engineering and Innovation, HQ SACT, Norfolk United States. ISBN 9789284501939

Yde, Iben 2016. "The Legal Implications of Weapons Systems with Autonomous Attack Capabilities - Towards an Understanding of the Changes in Human Influence on Target Selection and Engagement" Juridisk Institut. Aarhus Universitet, June 2016.