Normal Autonomous Accidents: What happens when killer robots fail?

Stephanie Carvin, Carleton University

Working Paper: March 2017

Abstract

Over the past decade there has been much written on lethal autonomous weapons systems (LAWS) commonly known as "killer robots". This includes legal, ethical and moral concerns as well as issues related to responsibility. And yet, for all of the discussion directed to concerns about what happens when something goes wrong, there is less attention paid as to how it will go wrong. The main difference between LAWS and our current weapons systems is the freedom they will have to make decisions on the battlefield, creating unique challenges for how we regulate such weapons. Not only is it potentially more difficult to verify the reliable behavior of LAWS, but the consequences of an accident could potentially be more severe.

As such, starting from the assumption that LAWS will be significantly more complex than our present weapons systems, this paper uses Charles Perrow's Normal Accident Theory (and its critics) to explore the concerns raised over LAWS through the lens of system failure. Focusing on failure provides new insights into problems related to risk mitigation strategies (such as weapons reviews) and responsibilities (chain of command). Moreover, it also points to preliminary steps the international community can take in regulating future weapons systems.

Introduction[1]

Over the last decade, there has been an increase in the scholarly attention directed towards lethal autonomous weapons systems (LAWS). Much of this literature is focused on interrogating the ethical, moral and legal issues that may arise in the future, and highlighting the key issues policy makers, practitioners and activists need to consider with in the present.[2] For example: What is the role of moral agency

[1] The author would like to thank XXX for their helpful comments.

[2] There are too many sources to cite here, but some examples include: Kenneth Anderson and Matthew Waxman, "Law and Ethics for Robot Soldiers", *Policy Review*, 1 December 2012. Available online: http://www.cfr.org/world/law-ethics-robot-soldiers/p29598; Kenneth Anderson and Matthew Waxman, "Law and Ethics for Autonomous Weapons Systems: Why a Ban Won't Work and How the Law of War Can", Hoover Institution, 9 April 2013. Available online: http://www.hoover.org/research/law-and-ethics-autonomous-weapon-systems-why-ban-wont-work-and-how-laws-war-can; Peter Asaro, "On banning autonomous weapon systems: human rights, automation and the dehumanization of lethal decision-making", *International Review of the Red* Cross, Vol. 94, No. 886, Summer 2012. pp. 687- 709; Chantal Grut, "The Challenge of Autonomous

Electronic copy available at: https://ssrn.com/abstract=3161446

in the decision to deploy weapons? Can autonomous systems apply the principles of distinction and proportionality? And who is responsible when an accident occurs?

The main difference between LAWS and non-LAWS is the freedom they will have to make decisions on the battlefield – LAWS are not reinventing the wheel, they are wheels that can spin themselves.[3] While our current system of international and domestic military weapons regulations are based on assumptions about control, authority, responsibility, proportionality and intent, the very nature of LAWS turns this on its head. While it is extremely unlikely that nations will deploy weapons they believe to be inherently unpredictable, it is the case that it will be much more difficult to verify the reliable behaviour of these machines. Moreover, the consequences of an accident with LAWS may also be more severe.

And yet, for all of the concern over what will happen if something goes wrong, there has not been given much attention to how it may go wrong. *How should countries and the international community begin to address the regulation of such weapons if there is much we do not know about how they may fail?*

While it may be challenging to do so with weapons that do not yet exist, we know enough about accidents and system failure to put forward ideas about how this will affect LAWS, their operations and how they are eventually regulated or banned in domestic or international law. Charles Perrow's "Normal Accident Theory" argues that two system characteristics, "interactive complexity and tight coupling… inevitably will produce an accident" in high-risk systems.[4] Perrow's theory suggests that accidents are an inescapable by-product of the very nature of these systems, regardless of organizational structure or chain-of-command and safety measures taken. Applying his theories to "high-risk systems" such as nuclear power plants,

---

Lethal Robotics to International Humanitarian Law", *Journal of Conflict and Security Law,* Vol. 18, No. 1, 2013. pp. 5-23; Human Rights Watch, *Mind the Gap: The Lack of Accountability for Killer Robots*, 2015. Available online: https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots; Chris Jenks, "False Rubicons, Moral Panic & Conceptual Cul-De-Sacs: Critiquing, Reframing the Call to Ban Lethal Autonomous Weapons", 2016. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2736407##; Valerie Morkevicius, "Tin Men: Ethics, Cybernetics and the Importance of Soul", *Journal of Military Ethics*, Vol. 13, No. 1, 2014. pp. 3-19; Gary E. Marchant et al. "International Governance of Autonomous Military Robots", *The Columbia Science and Technology Law Review*, Vol. XII, 2011. pp. 272-315; Heather M. Roff, "The Strategic Robot Problem: Lethal Autonomous Weapons in War", *Journal of Military Ethics*, Vol. 13, No. 3, 2014. pp. 211-227; Michael N. Schmitt and Jeffrey S. Thurnher, ""Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict", *Harvard National Security Journal*, Vol. 4, 2012-2013. pp. 231-281; Noel Sharkey, "Saying 'No!' to Lethal Autonomous Targeting", *Journal of Military Ethics*, Vol. 9, No. 4, 2010. pp. 369-383; Noel Sharkey, "The evitability of autonomous robot warfare", *International Review of the Red Cross*, Vol. 94, No. 886, Summer 2012; Robert Sparrow, "Killer Robots", *Journal of Applied Philosophy*, Vol. 24, no. 1, 2007. pp. 62-77.

[3] The author would like to thank XXXX for this observation.

[4] Charles Perrow, *Normal Accidents: Living with High-Risk Technologies*, Princeton: Princeton University Press, 1999. p. 5.

Electronic copy available at: https://ssrn.com/abstract=3161446

others such as Scott D. Sagan[5] and Scott A. Snook[6] have extended Normal Accident Theory to nuclear weapons and military systems respectively.

Therefore, this article starts from the premise that, the debate over LAWS can also be framed as one about accidents and "high-risk" technologies. As such, it draws on the insights of Perrow's classic *Normal Accidents* (and its critics) to view LAWS as tightly coupled, highly interactive complex systems. Crucially, with it emphasis on system failure, Normal Accident Theory allows us to produce a list of questions and grounded hypotheticals that provide the basis to generate corresponding insights into the debate over LAWS.[7] First, regardless of safety mechanisms, coding precautions and standard operating procedures, and legal reviews, accidents will inevitably occur. While the so-call "fog of war" will inevitably play a role in this, accidents will be the result of inevitable, unanticipated interaction of system components. Second, when these accidents occur it will not be possible to determine why a failure occurred and very challenging to assign responsibility for them.

As such, after briefly discussing the problems raised by LAWS and failure in automated systems, this paper will use the debate between Normal Accident Theory and its critics to highlight some of the most salient issues about the failure and regulation of these systems for the international community. The final part of this paper will argue that based on the findings above, our current frameworks are inadequate for thinking about LAWS and this has serious implications for the law of armed conflict and thinking about regulation or bans on such weapons. While the principles of customary and international treaty law codified in the late 19th and early 20th centuries have been successfully applied to conventional weapons, including unmanned aerial vehicles (UAVs), Normal Accident Theory suggests that for the first time international society may be dealing with weapons that challenge this framework. This is especially the case as we move away from more "linear" systems featuring direct sequences of command and control, to systems that are not only non-linear, but function at a level of unique complexity that is significantly greater than anything currently deployed in military operations.[8] This raises the

---

[5] Scott D. Sagan, *The Limits of Safety: Organizations, Accidents and Nuclear Weapons*, Princeton: Princeton University Press, 1993.

[6] Scott A. Snook, *Friendly Fire: The Accidental Shootdown of U.S. Blackhawks over Northern Iraq*, Princeton: Princeton University Press, 2000.

[7] Paul Scharre discusses Normal Accident Theory and LAWS in report "Autonomous Weapons and Operational Risk" Washington D.C.: Centre for New American Security, February 2016. Available online: http://www.cnas.org/sites/default/files/publications-pdf/CNAS_Autonomous-weapons-operational-risk.pdf. The present paper, however, seeks to not only note how the theory applies to the notion of operational risk, but also how such a perspective can highlight some of the possible difficulties with the systems, certain policy proposals and suggest certain pathways forward as the international community continues to grapple with these issues.

[8] This is especially the case given recent advances in the development of neural networks which learn from massive amounts of data and operate in a very different way from rule-based systems. While there is not the space to engage in a lengthy discussion of neural networks, the general idea is to make artificial systems that mimic the human brain. Decision making becomes decentralized over thousands of little equations (or "neurons") that take the data, process it and pass it onto another

distinct possibility that such weapons may fail more often or fail in ways that cannot be anticipated. As such, even if principles are "codified" into autonomous systems, the likelihood of failure, and challenges in determining responsibility will require another way of thinking about accountability under the laws of armed conflict. As such, the paper will present some preliminary steps the international community should take going forward. This includes generating ways to think about responsibility in ways that go across a number of actors, developing best practices, bringing private industry to the table with government and humanitarian actors and continued diplomatic efforts to address ongoing technological developments.

All By Myself: Why Autonomy?

Despite an increase in the amount of autonomy and autonomous function ins our every day lives, the concept is challenging to define. As Paul Scharre notes, it "is not some point we arrive at in the future. Autonomy is a characteristic that will be increasingly incorporated into different functions on military systems."[9] A similar view is shared by Bradshaw et al, who note that autonomy is not a "widget", but "an idealized characterization of observed or anticipated interactions between the machine, the work to be accomplished and the situation."[10] Indeed, what is "autonomous" will vary, depending on the task at hand, where it is taking place, how something is linked to other systems. As such, it is not appropriate to think about autonomy as a spectrum, but rather to focus on the autonomous function within systems – or *operationally-relevant autonomy*, the "sufficient autonomy to get the job done."[11] Building on this idea, this paper will use the definition of LAWS offered by Paul Scharre and Michael Horowitz, that is: "weapons systems, that, once activated, [are] intended to select and engage targets where a human has not decided those specific targets are to be engaged."[12]

---

layer of thousands of little equations. David Gershgorn, "Fooling the Machine", *Popular Science*, 30 March 2016. Available online: http://www.popsci.com/byzantine-science-deceiving-artificial-intelligence

[9] Paul Scharre, "Between a Roomba and a Terminator: What is Autonomy?", *War on the Rocks,* 18 February 2015. Available online: http://warontherocks.com/2015/02/between-a-roomba-and-a-terminator-what-is-autonomy/

[10] Jeffrey Bradshaw et al, "The Seven Deadly Myths of "Autonomous Systems"", *IEEE Intelligent Systems*, May/June 2013. pp. 2-9. p. 4.

[11] Scharre, "Between a Roomba and a Terminator".

[12] Paul Scharre and Michael C. Horowitz, "An Introduction to Autonomy in Weapon Systems" Centre for New American Security Working Paper, February 2015. Available online: http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf. Scharre notes that rather than searching in vain for a unified framework of "levels of autonomy," that "a more fruitful direction is to think of autonomy as having three main axes, or dimensions, along which a system can vary. These are the human-machine command-and-control relationship, the complexity of the machine, and the type of decision being made. In this sense, it is important to think less about whether we will get to "full autonomy" as there is not a spectrum along which autonomy moves. Instead a better framework is "to ask which task are done by a person and which by a machine." Or thinking about the "autonomous functions" of systems, rather than characterizing an entire vehicle or system as

Although the topic seems novel, states have been using weapons with autonomous components for decades. By 2015 over 30 countries were known to have defensive systems with human-supervised autonomous modes.[13] Within these systems, autonomy is used for military tasks such as identifying, prioritizing and cueing targets, timing of when to fire and detonate, and maneuvering and homing in on targets.[14]

And yet, it is important to note just how complex these systems will be relative to the military systems currently in use. A helpful starting point is the difference between linear and non-linear systems. Linear control systems are those in which the relationship between the robot and the controller can be described by linear differential equations. In other words, what comes out of the system is directly related to what goes into the system – a situation often described as the superposition principle.[15] Linear systems and their interactions are deterministic in that they are straightforward, planned, and visible. This renders them simple to comprehend because we can easily know what the system did in the past and we can predict what it will do in the future.

Non-linear systems are systems that are *not* governed by the superposition principle and are more likely to generate irregular patterns. This often occurs in systems with feedback loops, branching paths, or feature unpredictable jumps between different system components. For example, in a non-linear system the output generated may be more or less than the sum of what went into the system.

Humans benefit from innate skills to help them deal with a complex and non-linear world. This includes an ability to make common-sense decisions (like everyday physics and psychology) as well as the ability to understand and learn entirely new abstract concepts and build models.[16] We are assisted in this because biological systems are non-linear and are able to adapt to their environment, changing their control systems accordingly. Take the case of a person who shows up to work to find their colleagues standing outside in the rain. If that person asked "what are you doing?" it would be strange if he or she answered "standing in the rain". Instead, if he or she replied "Fire alarm", it would be possible to instantly figure out what was happening based on prior experience, common sense and shared understandings.[17]

"autonomous". Scharre's concise and useful description of autonomy can be found in his article, "Between a Roomba and Terminator"

[13] Scharre and Horowitz, "An Introduction", p. 3.

[14] Scharre and Horowitz, "An Introduction", p. 3. In their report, the authors provide an appendix of "Selected Examples of Human-Supervised Autonomous Weapons Systems Currently in Use". See Appendix B. pp. 21-23.

[15] George A Bekey, *Autonomous Robots: From Biological Inspiration to Implementation and Control*, Cambridge, MA: The MIT Press, 2005. p. 9. Put simply, in linear systems, if the input is doubled, the output will also be twice as large.

[16] Murray Shanahan, *The Technological Singularity*, Cambridge, MA: MIT Press, 2015. pp. 55-58

[17] Example borrowed from Shanahan, *Technological Singularity*, p. 7.

This is different from our present engineering systems (including most weapons and platforms) that tend to be fixed and non-adaptive.[18] At present, the bulk of interactions within engineered systems are linear. When autonomous systems encounter a diverse, biological, non-linear world, they will have to deal with irregularity, learn, adapt and handle uncertainty. This will include having to visualize situations, predicting the consequences of actions in circumstances they have not previously encountered – something that will take a considerable feat of engineering.

In brief, the kind of systems or processing power required for dealing with uncertainty does not yet exist. On the challenges of generating artificial general intelligence (AGI) that could potentially reach human-level, Shanahan notes:

> The engineering challenge here is not merely to achieve the required number of FLOPS (floating point operations per second) but to do so in a small volume and with low power consumption. The average human brain (male) occupies a mere 1250 cm^3 and consumes just 20 W [of power]. By contrast, the Tianhe-2, the world's most powerful supercomputer in 2013 consumes 24 MW and is housed in a complex occupying 720m^2. Yet it still has only a fraction of the computing power needed to simulate a human brain under even the most conservative assumptions.[19]

Nevertheless, it is clear that science is working towards ever faster and significantly more complex computers. For example, scientists in several countries are working towards developing computers that can conduct a petaflop, or a quadrillion floating point operations per second – or more.[20] These will be ideal for modeling systems such as monitoring pollution, but also understanding environments such as the battlefield. Further, recent developments in neural networks and deep learning mean that tasks that were onerous for computer systems even 6 years ago (such a as face and pattern recognition) are now far more advanced.[21]

---

[18] Bekey, *Autonomous Robots*, p. 10.
[19] Shanahan, *Technological Singularity*, p. 32. It should be noted that Shanahan is relatively optimistic about finding biologically inspired ways around this problem. See Chapter 2 "Whole Brain Emulation", pp. 15-50.
[20] See Michael Clooney, "Beyond the petaflop: DARPA wants quintillion-speed computers", *Network World*, 23 June 2010. Available online: http://www.networkworld.com/article/2231123/security/beyond-the-petaflop--darpa-wants-quintillion-speed-computers.html
[21] See for example, Cade Metz, "In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go", *Wired*, 27 January 2016. Available online: https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/; Michael Thomsen, "Microsoft's Deep Learning Project Outperforms Humans In Image Recognition", Forbes, 19 February 2015. https://www.forbes.com/sites/michaelthomsen/2015/02/19/microsofts-deep-learning-project-outperforms-humans-in-image-recognition/#49ef8484740b

While there is no consensus over when practical supercomputers or AGI may be possible,[22] or the kind of processing power required for dealing with uncertainty in a feasible (or at least compact) way, there are some key insights from this brief overview for this article. In particular, it is not yet clear that LAWS will require full AGI, human-level intelligence, or supercomputers for all potential operational environments. Battlefield complexity is relative to the operational environment; an empty desert battlefield or ocean surface will be far less complex than that found in an urban environment or jungle. However, operating in these latter environments will take a great deal of complexity – and it is not yet clear how feasible this will be.

Additionally, there is near consensus among the academics, international organizations, governments and NGOs working in this area that the LAWS viewed as problematic are different from the weapons used by the militaries today.[23] Instead, what is of primary concern are weapons that will have no "human in the loop" or no "meaningful human control."[24] These weapons do not yet exist (and given the complexity required as discussed above, it is not clear they can actually exist[25]) although for many involved in campaigns against "killer robots", there is concern that states are actively developing them.[26]

Indeed, it is almost certain that states will be moving towards using weapon systems and platforms with increasing levels of autonomy for a number of operational reasons. First, greater autonomy means fewer personnel will be required to operate the machinery. Scharre notes "Unmanned systems can be designed with performance characteristics and risk profiles that for some missions simply would be not feasible or acceptable for manned vehicles."[27] It allows states to take on more hazardous missions as unmanned systems decouple force protection from combat performance.[28]

---

[22] For example the a report by President Barack Obama's National Science and Technology Council Committee on Technology concluded that human-level artificial intelligence will not be achieved for decades. Experts surveyed for the study offered dates between 2030 and centuries from now. "Preparing for the Future of Artificial Intelligence", October 2016. P.7 Available online: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

[23] Michael C. Horowitz, "The Ethics and Morality of Robotic Warfare: Assessing the Debate Over Autonomous Weapons", *Daedalus*, *FORTHCOMING.*

[24] See, for example, Heather M. Roff and Richard Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons", Article 36, April 2016. Available online: http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf; Human Rights Watch, "Mind the Gap".

[25] See, for example, the points raised by Bradshaw, et al., who note that autonomy is a "characteristic" that is contextual depending on the task. Further they note that autonomous systems are likely to require the involvement of people in roles where expertise is a must and what is needed is not more autonomy, but breakthroughs in human-machine teamwork that would not replace people, but assist them." pp. 5-6

[26] See the Campaign to Stop Killer Robots, https://www.stopkillerrobots.org/; International Committee for Robot Arms Control, http://icrac.net/

[27] Paul Scharre, "Why Unmanned?", *Joint Forces Quarterly*, No. 61, 2nd Quarter, 2011. pp. 89-93 p. 89.

[28] Scharre, "Why Unmanned", p. 90.

Second, Michael N. Schmitt and Jeffrey S. Thurnher note that the advantage of autonomy over systems tethered to a human operator is that they are less vulnerable to satellite communications jamming and cyber-attack. Autonomous weapons will be able to continue to operate even if a communications link is severed.[29]

Finally, Schmitt and Thurnher suggest that "Future combat may therefore occur at such a high tempo that human operators will simply be unable to keep up."[30] Already, automation is required in many advanced military systems such as missile evasion or where pilots would otherwise be overwhelmed by the amount of information they are receiving when they are already under stress.[31] Autonomy allows for greater speed and levels of complexity that will be required for future capabilities and it is likely that militaries around the world will be looking to incorporate these capabilities into their future weapon systems.

To Err is Robotic…

To anyone who has experienced the Windows "blue screen of death", or the relentless "spinning beach ball" on their MacBook, it is obvious that computerized systems are prone to failure. But although these may be annoying, they are not weaponized platforms, capable of misfiring or attacking the wrong target. Indeed, LAWS, significantly more complex than everyday computers and our present weapons platforms, will likely require millions upon millions of computations per second. They will be designed to operated in hostile environments, facing an array of challenges, including telling friend from foe and establishing proportionality. However, given all of this unique complexity, an important question arises: what will happen when these systems fail?[32]

Schmitt and Thurner argue that LAWS will not "go rogue" and that "the prospect of them 'taking a life of their own' is a fantastical Hollywood invention."[33] In other words, we should not worry ourselves with *Terminator* scenarios just yet – indeed, states, such as the UK, have explicitly stated they are not interested in building

[29] Schmitt and Thurnher, "Out of the Loop", 238.
[30] Schmitt and Thurnher, "Out of the Loop", 238.
[31] Stephen E. White, "Brave New World: Neurowarfare and the Limits of International Humanitarian Law", *Cornell International Law Journal,* Vol. 41, No. 1, 2008. pp. 177-210.
[32] The failure discussed in this paper relates to system malfunction. However, there are clearly other obvious ways autonomous systems could fail. For example, it might be deployed in the wrong way, there may be unexpected interactions with or changes in the environment after the system is deployed, etc. These, however, are accidents that could happen to any weapon that presently exists and are not the focus of this paper. More problematic are those actions taken by hostile actors to fool or "spoof" AI. These are, however, not "normal accidents" *per se* and therefore also fall outside of the scope of this paper. However, they raise interesting practical and scholarly questions, especially related to what has traditionally been called "ruses of war" in the law of armed conflict (Article 37(2), Additional Protocol I to the 1949 Geneva Convention).
[33] Schmitt and Thurnher, "Out of the Loop", p. 242.

movie-style "killer robots" and that humans will always remain involved in targeting decisions.[34]

While these points may be true, they are also insufficient. In particular, they fail to consider that some failures are more straightforward than others. When a missile misfires, we can largely predict what it will do and the impact it will have. However, when it comes to autonomy and AI, dealing with failure is extremely challenging for at least four important reasons. First, systems are not always good at communicating what is going on when something is going wrong. Bradshaw et all argue, "humans and machines working together frequently encounter potentially debilitating problems relating to insufficient observability or understandability…". A key problem with automation is that:

> …it fails to communicate effectively those things that would allow humans to work interdependently with it – signals that allow operators to predict, control, understand, and anticipate what the machine is or will be doing. As anyone who has wrestled with automation can attest, there's nothing worse than a so-called smart machine that can't tell you what it's doing, why it's doing something, or when it will finish."[35]

Second, and potentially worse, malfunctioning systems can provide misleading information, confounding the operator further. Lamport, et al note that, "A failed component may exhibit a type of behavior that is often overlooked – namely, sending conflicting information to different parts of the system… "it is often assumed that a computer may fail to respond but will never respond incorrectly."[36]

This relates to a third problem highlighted by Amodei et al, described as a "scalable oversight" problem. Given the large number of functions the system is performing, it is extremely difficult to know if all of its aspects are functioning correctly. We may only be able to evaluate a certain number of functions at any given time, and this raises the dilemma of bad extrapolations from limited data. In other words, given limited access to the full set of functions the system is performing, how can we be sure the system is performing as it should?[37]

Finally, systems designed for one kind of scenario may not operate well outside of this set of parameters. Horowitz and Scharre note, "the same features that make [LAWS] reliable – the fact that they follow their programming precisely every time – can make them brittle when used outside of their intending operating

---

[34] Ned Simons, "Britain Not Building 'Killer Robots', Minister Insists", *HuffingtonPost UK*, 18 June 2013. Available online: http://www.huffingtonpost.co.uk/2013/06/18/killer-robots-britain_n_3459463.html

[35] Jeffrey Bradshaw et al. p. 3.

[36] Leslie Lamport, Robert Shostak and Marshall Pease, "The Byzantine Generals Problem", ACM Transactions on Programming Languages and Systems, Vol. 4, No. 3, July 1982. pp. 382-401.

[37] Dario Amodei et al, "Concrete Problems in AI Safety", Available online: https://arxiv.org/abs/1606.06565

environment."[38] What happens when LAWS are asked to operate outside of the box and in the fog of war? Amodel et al describe a similar concern as "robustness to distributional shift" – how can we ensure that LAWS behaves robustly when it is in an environment different from its training or testing environment?[39]

To be clear, there are many other types of failure that do not arise from malfunction. For example, systems can be deployed in the wrong way, or be tricked by deliberate attempts by adversaries to "hack" and fool (also known as "spoofing") the system.[40] However, the reason these four kinds of failure are important for the argument here is that they are *latent* failures that are inherent in the system and do not require external action to manifest. As such, in a crisis or emergency, they will constitute the failures that are hardest to recognize and solve.

To deal with these risks, the US Department of Defense has already directed all autonomous systems "be readily understandable to trained operators" and "provide traceable feedback on system status".[41] However, given the problems outlined above, we cannot expect that operators will know why something is going wrong in time to stop an accident from happening. So if transparency is so difficult, how could we ever know if such weapons are compatible with the laws of war?

Essentially, this is the key dilemma – we know that states are designing complex weapons that will occasionally fail in unpredictable ways. Further, given the nature of these systems where decisions are being made using millions upon millions of computations in multifaceted environments, across weapons platforms that may be overseen by many individuals, how will we be able to investigate accidents? Is such a system really compatible with our current understanding of command responsibility, and the rules governing the use of force?

All Systems Go? Normal Accidents

Given its focus on accidents in "high risk" systems, Normal Accident Theory and its critics provides an excellent framework in which to consider LAWS, accidents and the implications of failure in these systems. The origins of this approach rests with Charles Perrow's 1984 book *Normal Accidents,* which analyzes several areas of industry and innovation that had experienced accidents in the 1960s and 1970s: nuclear power, the chemical industry, aircraft, space exploration, etc. (The 1999 updated version also briefly assesses Chernobyl and the Y2K computer bug.)

---

[38] Michael C. Horowitz and Paul Scharre, "The Morality of Robotic War", *New York Times*, 27 May 2016. Available online: http://www.nytimes.com/2015/05/27/opinion/the-morality-of-robotic-war.html

[39] Amode et al, "Concrete Problems in AI Safety", p. 3.

[40] The author is grateful to ____ for raising this point.

[41] Department of Defense Directive Number 3000.09, "Autonomy in Weapons Systems", 21 November 2012. Available online: http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf

What is unique about Perrow's approach is that it was the first to focus on the characteristics of systems, rather than mistakes of operators and owners. He finds that accidents are often attributed to the actions of individuals, when in reality, they are often the product of *systemic* characteristics. In Perrow's account, systems are comprised of four levels of increasing aggregation: (1) *parts* that make up (2) *units*, which form (3) *sub-systems* that create the (4) *system*. As discussed above, when the interactions within and between these parts are straightforward, planned and visible, they are *linear* in nature. These interactions reflect the way the system is designed to run and any operator running the system should be able to recognize them as normal.[42] However, unpredictable non-linear interactions in complex systems creates circumstances where an unanticipated connection between two unrelated sub-systems may occur. As such, these interactions are less comprehensible (especially to operators who will not be able to understand what has occurred) and more difficult to understand.

"System accidents" occur when there are multiple failures in largely independent subsystems that have interacted in unanticipated ways. Perrow argues that accidents are inevitable in systems that share two characteristics. The first of these is *interactive complexity*, where system components are highly interactive with one another as described above. Second, when these systems are *tightly coupled* - where processes in a system happen very fast, affecting the ability for them to be shut off, where failed parts cannot be isolated from other parts, or where there is no other way to keep the production going safely.[43] According to Perrow, the very nature of these systems results in a situation where parts in these systems will inevitably interact in an unanticipated way – or in a way that could not have been anticipated by the systems designers, constructors, supervisors or operators so it could have been prevented.

Already we can see similarities with the concerns over autonomy listed above. Complex systems have elaborate controls because they make life easier for operators, saving steps or time, "not because there is necessarily more machinery to control, but because components must interact in more than linear, sequential ways, and therefore may interact in unexpected ways."[44] As systems are growing more and more complex, and it is impossible for designers, managers and operators of these systems to anticipate the almost limitless ways that parts, units, and systems can interact. Therefore, for Perrow, system accidents are rare, but inevitable:

> Since nothing is perfect – neither designs, equipment, operating procedures, operators, materials, and supplies, nor the environment – there will be failures. If the complex interactions defeat designed-in safety devices or go around them, there will be failures that are unexpected and incomprehensible. If the system is also tightly coupled,

[42] Perrow, *Normal Accidents*, pp. 75-78.
[43] Perrow, *Normal Accidents*, p. 4.
[44] Perrow, *Normal Accidents*, p. 83.

> leaving little time for recovery from failure, little slack in resources or fortuitous safety devices, then the failure cannot be limited to parts or units, but will bring down subsystems or systems…
> Much can be done to make these systems somewhat safer, but accidents cannot be entirely avoided. Quality control, operator training, design experience, and environmental controls will help, but will not be sufficient.[45]

While complexity poses risks to systems such as airplanes and our networks of highways, there is considerable danger when systems that are interactively complex and tightly coupled are applied to in high-risk areas such as nuclear power and nuclear weapons.[46] In particular, Perrow expresses concerns about "transformation systems" that transform raw materials rather than fabricate or assemble them. This includes recombinant DNA technology, chemical plants, nuclear power production, nuclear weapons, and some aspects of space missions (particularly those with nuclear elements.)[47] These have the potential to generate catastrophes (defined as and accident that kills more than 100 people with one blow). As such, when building such systems, it is incumbent on society to extend "the logic of interacting failures to the consequences of system failure – an accident", when this potential exists.

The Normal Autonomous Accident?

Since it was published in 1984, there has been considerable reaction to Normal Accident Theory, with some authors seeking to extend the theory, such as Scott Sagan and Scott Snook (among others) into a more explicitly organizational, managerial, social and political formulation.[48] However, there has also been a fair amount of criticism levied at Perrow's argument. While there is not the room to enter into a detailed discussion of all of these critiques, highlighting a few helps us to think through the implications of Normal Accident Theory for LAWS.[49]

---

[45] Perrow, *Normal Accidents*, p. 330.
[46] Perrow, *Normal Accidents*, p. 304.
[47] Perrow, *Normal Accidents*, p. 85.
[48] Jean-Christophe Le Coze, "1984-2014. Normal Accidents. Was Charles Perrow Right for the Wrong Reasons?", *Journal of Contingencies and Crisis Management*, Vol. 23, No. 4, December 2015. pp. 275-286. p. 284. See also Sagan, *The Limits of Safety*; Snook, *Friendly Fire*.
[49] It is important to acknowledge that Perrow's original work focuses more on civilian than military technology. Although he has occasionally addressed some military technologies throughout his career, (on Network Centred Warfare) Perrow makes some errors in his models that suggest that he may not have been thinking through military models as much as he had civilian ones. For example, as Snook notes in a key chart in his book, Perrow considers certain military actions, often part of one larger military system, as separate activities. See Snook, *Friendly Fire*, pp. 14-15. For Perrow's work on Network Centered Warfare see Charles Perrow, "Difficulties with Network Centric Warfare" in Jacques S. Gansler, Hans Binnendijk, *Information Assurance: Trends in Vulnerabilities, Threats and Technologies*, Washington DC: National Defense University, 2004. pp. 139-146. This article largely focuses on whether NCW could actually deliver on its promises ("the dreams of Admiral Cebrowski, considered by many to be the father of the NCW concept, and those of others likely will not be realized in the next couple of decades") and that it may actually make the military counterproductive (leading to micromanagement).

Unfalsifiablity

A first and important criticism that may be raised is that Normal Accident Theory is unfalsifiable.[50] Antti Silvast and Ilan Kelman note that Normal Accident Theory rests on two assumptions. First, no past record of the absence of a normal accident excludes the possibility of a future normal accident. (Or, as they phrase it "The standard scientific mantra expressing this assumption is that 'absence of evidence is not evidence of absence.")[51] This is also a complaint of Hopkins who argues that this assumption makes Normal Accident Theory impossible to test.[52] Second, the "analysis of a normal accident is always relative to a selected definition of the system. Hence, a system can always be defined so that there was or was not a Normal Accident.[53] Indeed, several commentators and critics share similar concerns that Perrow does not define many of the terms core Normal Accident Theory.[54]

That "system" or "complexity" may be defined and re-defined is an important critique of Normal Accident Theory, as it highlights the risk that it may become incoherent if such a key term is so loosely defined. Yet, Perrow wants to keep the definition flexible so that it may change according to circumstances as well as scholarly interest. As Silvast and Kelman note, "a system is something that one chooses to analyse to highlight a particular kind of damage or disruption."[55] In this way, a more strict approach could do more harm than good: it may prevent taking a useful perspective to larger issues, such as Sagan's work on the command and control of nuclear weapons or Snook's work on "friendly fire" accidents.[56] Indeed, as will be seen below, the debate over LAWS benefits from when we take flexible view of what constitutes a system, whether we focus on the platform, or the organization that controls it. As such, there are advantages to leaving "system" as a contestable concept that allows for a broad debate over accidents. Therefore, even if Normal Accidents Theory is a common sense "truism"[57], rather than an iron-clad law, it still

[50] Eugene A. Rosa, "Celebrating a Citation Classic – and More", *Organization & Environment*, Vol. 18, No. 2, June 2005. Pp. 229-234.
[51] Antti Silvast and Kelman, "Is the Normal Accidents perspective falsifiable?", *Disaster Prevention and Management,* Vol. 22, No. 1, 2013. pp. 7-16. p. 9.
[52] Andrew Hopkins, "The limits of normal accident theory", *Safety Science* Vol. 32, 1999. pp. 93-102. pp. 97.
[53] Silvast and Kelman, "Is the Normal Accident perspective falsifiable?", p. 9.
[54] See, for example, Shrivastava et al. "Normal Accident Theory versus High Reliability Theory: A resolution and call for an open systems view of accidents", *Human Relations*, Vol. 62, No. 9, 2009. pp. 1357–1390. p. 1359. Hopkins notes the "absence of clear criteria for measuring complexity and coupling", "The limits of normal accident theory", p. 96.
[55] Silvast and Kelman, "Is the Normal Accident perspective falsifiable?", p. 12. Perrow, *Normal Accidents,* pp. 63-66.
[56] Sagan, *The Limits of Safety*, Snook, *Friendly Fire*.
[57] Silvast and Kelman, "Is the Normal Accident perspective falsifiable?", p. 8 and p. 14.

provides a useful perspective from which to explore issues related to high-risk technologies, including LAWS.[58]

Technologically Determinist

A second criticism of Perrow's work is that he focuses on technology to the detriment of other factors, including the role of actors.[59] For example, Andrew Hopkins argues an "unashamedly technological determinist argument" constitutes a fatal weakness for Perrow's theory.[60] Instead, he subscribes to High-Reliability Theory which maintains that certain organizations, despite their complexity and tight coupling, were nevertheless able to achieve outstanding safety records.[61] These "high reliability organizations" follow certain practices, such as subjecting their personnel to intense training and socialization to make sure they are able to respond to hazardous contingencies swiftly, the use of redundancy to back up failing parts and persons in their organization, etc. As such, these organizations appear to be failure free, because they put a high premium on reliability (where reliability is the ability to maintain and execute error-free operations) since their operating environments seldom offer them a second chance.[62] More critically, Hopkins argues that Normal Accident Theory is not useful because it only applies to a "very small, and furthermore ill-defined subset of disasters or near disasters."[63] As such, Normal Accident Theory "has nothing to say about many of the most publicized disasters of our time."[64]

---

[58] Importantly, many scholars have noted that although Normal Accident Theory and High Reliability Theory appear diametrically opposed, that there is a great deal upon which they agree. Indeed, one of the key proponents of High-Reliability Theory, La Porte, argues that the two perspectives are complementary. Todd R. La Porte, "A strawman speaks up: Comments on *The Limits of Safety*", *Journal of Contingencies and Crisis Management*, Vol. 2, No. 4, 1994. pp. 207-211. p. 211. Rijpma argues that Perrow and Hopkins have a very similar view of "safety cultures" in their argument, and their failures. Rijpma, "From Deadlock to Dead End", p. 42. Shrivastava et al, argue that Normal Accident Theory and High Reliability Theory are not incommensurate, as they "refer to the same phenomenon, but at different time frames.", "Normal Accident Theory", pp. 1374.

[59] Le Coze, "1984-2014", p. 276. See also Jos. A. Rijpma, "From Deadlock to Dead End: The Normal-Accidents-High Reliability Debate Revisited", *Journal of Contingencies and Crisis Management*, Vol. 11, No. 1, March 2003. pp. 37-45. p. 38.

[60] Andrew Hopkins, "Was Three Mile Island a 'Normal Accident'?" *Journal of Contingencies and Crisis Management*, Vol. 9, No. 2, June 2001. pp. 65-72. p. 65.

[61] Todd R. La Porte and Paula M. Consolini, "Working in practice but not in theory: theoretical challenges of "high reliability organizations", *Journal of Public Administration Research and Theory*, Vol 1. No. 1, 1991. pp. 19-47.

[62] Samir Shrivastava, Karan Sonpar, and Federica Pazzaglia ,"Normal Accident Theory versus High Reliability Theory: A resolution and call for an open systems view of accidents", *Human Relations*, Vol. 62. No. 9, 2009. pp. 1357-1390. p. 1363 and Rijpma, "From Deadlock", p. 41.

[63] Hopkins, "limits of normal accident theory", p. 101.

[64] Hopkins, "limits of normal accident theory", p. 95. Instead, Hopkins draws on Disaster Incubation Theory which argues that disasters are caused by failures of foresight when, during long incubation periods, signals about impending danger are either ignored or misunderstood until it is too late. Combined with an overly-optimistic view of safety, poor managerial practices, accidents occur, although it may take many years for the latent danger to actually manifest into a disaster. See Rijpma,

Importantly, Perrow himself agrees with this view – normal accidents are rare events. Further, they must be distinguished from accidents that are the fault of bad policies and mismanagement. For example, Perrow argues that neither Chernobyl, nor the Bhopal disaster nor the Challenger and Exxon Valdez accidents meet his definition of a normal accident: "They are alarmingly banal examples of organizational elites not trying very hard at all and are what I call 'component failure' accidents.[65] In other words, they are not the product of highly interactive complexity, but preventable, sloppy management practices. While Hopkins sees this as a flaw in Perrow's argument, Perrow himself seeks to emphasize the rarity of normal accidents (and the idea that they are distinct events), which makes them more dangerous: rare, unanticipated and in high-risk technological areas, the potential for catastrophe is often out of mind, but all too real.[66]

Weapons Review

When it comes to LAWS, a High Reliability Theory practice might be considered a weapons review.[67] Article 36 of Protocol I to the 1949 Geneva Conventions requires that weapons, means and methods of warfare be evaluated to ensure they comply with the laws of war.[68] In theory, this would ensure that LAWS are discriminate (can distinguish between civilians, civilian objects and legitimate targets) and do not cause unnecessary suffering. It should also ensure that weapons disproportionately prone to error would never be deployed on the battlefield. However, even those who argue that LAWS likely would be compliant with international law note that "Given the technological advances likely to be embedded in autonomous weapons, this straightforward task may be challenging."[69]

"From Deadlock", p. 40 and the work of Brian A. Turner and Nick Pidgeon, *Man-Made Disasters* (second ed.), Oxford: Butterworth-Heinemann, 1997.

[65] Charles Perrow, "The Limits of Safety: The Enhancement of a Theory of Accidents", *Journal of Contingencies and Crisis Management*, Vol. 2, No. 4, 1994. p. 212-220. p. 218. See also Charles Perrow, "A Personal Note on *Normal Accidents*", *Organization & Environment*, Vol. 17, No. 1, March 2004. pp. 9-14. p. 10. Perrow has explicitly rejected attempts to apply Normal Accident Theory to these accidents, such as Diane Vaughn, *The Challenger launch Decision: Risky Technology, Culture and Deviance at NASA*, Chicago: University of Chicago Press, 1996.

[66] Further, Le Coze argues that criticisms of technological determinism in Perrow's argument fail to take into consideration the extended and alternate versions of Normal Accident Theory articulated by Perrow and several other theorists, which has helped Normal Accident Theory evolve into a more sociological approach (as discussed above). These critics fail to consider how *Normal Accidents* should be situated within Perrow's broader work on organizations and organizational failureLe Coze, "1984-2014", pp. 276-278.

[67] To be clear, this would only be one step in an overall HRT-style program.

[68] The text of article 36 states: In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.

[69] Schmitt and Thurnher, "Out of the Loop", p. 273. The article provides a useful overview of Article 36 reviews: pp. 271-276.

Legal reviews also raise another challenge – they are based on the specific *intent* for the weapon, not all possible uses. (For example, guns are evaluated on their basis to fire a bullet, not on their potential to be used as an object to beat a prisoner.) Schmitt and Thurnher argue that because usage is "contextual, it is generally inappropriate to make *ex ante* judgments [about the rules]".[70] However, even evaluating the intent of such a weapon will prove to be difficult. As Horowitz and Scharre argue, "After you fire a bullet, you can't take it back, but its trajectory is predictable. They key is to ensure that future weapons that behave like self-steering bullets do not run amok."[71] In essence, the task is not understanding how a weapon fires and impacts upon its target, but how that weapon *decides* to do so – a situation that is entirely context specific. While such a weapon could be put through multiple scenarios as part of its testing, it would be impossible to test for all possible contingencies in a way that it would make it possible for a commander to take "all feasible precautions" as required by law.

To provide a much less complicated analogy than war, an April 2016 study of autonomous cars found that autonomous vehicles, operating in a somewhat chaotic and yet relatively controlled traffic grid with a clear set of rules, could not be test-driven enough miles to demonstrate their safety.[72] While no weapon is perfect, and that a certain amount of unpredictability is to be expected as weapons are tried in actual battle conditions for the first time, the very nature of LAWS suggests that a weapons review would provide little guidance as to their limits and safety. Therefore, as Schmitt and Thurnher note, it is doubtful that reviews will really provide impediment to the development of LAWS.[73]

Finally, less than 20 states have Article 36 weapons reviews, and even then testing by a state would only cover their own systems.[74] As such, it will be hard to predict what might happen when two autonomous systems from different countries encounter one another, even by accident. Normal accidents occur when two components interact with one another in ways never intended by their creators, owners or operators, we should very well expect serious challenges when LAWS begin to regularly encounter one another.[75]

---

[70] Schmitt and Thurnher, "Out of the Loop", p. 274.
[71] Horowitz and Scharre, "The Morality of Robotic War".
[72] Nidhi Kalra, Susan M. Paddock, *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?*, Washington DC: RAND, 16 April 2016. Available online: http://www.rand.org/pubs/research_reports/RR1478.html
[73] Schmitt and Thurnher, "Out of the Loop", p. 276.
[74] SIPRI, "Implementing Article 36 weapon reviews in the light of increasing autonomy in weapon systems", 11 November 2015. Available online: https://www.sipri.org/media/press-release/2015/implementing-article-36-weapon-reviews-light-increasing-autonomy-weapon-systems.
[75] Horowitz and Scharre suggests a "flash war" like the 2010 "flash crash" could occur. Horowitz and Scharre, "The Morality of Robotic War". The author believes this may be overstated, but that unpredictable results from unanticipated interactions in tense situations could certainly produce negative results. A stronger point is made by Roff who notes that that "international or allied joint forces activities are plagued by operational difficulties due to the joining of different militaries.

Organization Theory

As noted above, High-Reliability Theorists argue that good practices will be able to mitigate safety issues, allowing for the safe use of LAWS. Beyond weapons reviews, this may include the practices of "High Reliability Organizations" such as aircraft carriers.[76]

Perrow specifically discusses weapons in *Normal Accidents* in his chapter on "Exotics: Space, Weapons and DNA". Interestingly, despite noting several near-misses, and the level of danger associated with military technology, Perrow seems relatively optimistic when it comes to the command and control of nuclear weapons at NORAD:

> "the detection system – NORAD and the sensing systems- is moderately complexly interactive. Linearity is introduced because the subsystems are independent of each other, not proximate spatially, nor subject to many common-mode components (though there are some.) Some loose coupling is present because recovery is possible from some low-level failures… Some loose coupling is built into the system because they can cover contingencies designers did not think of.)[77]

Although Perrow notes that it is the job of the "enemy (supremely clever and resourceful, it is assumed)" to trick the other side, Perrow concludes that "the early warning system appears to be moderately complex and coupled, but not disastrously so".[78] Indeed, Perrow attributes a strict "military" model applied to nuclear weapons that has a high degree of organizational control over its members as an important factor in enhancing safety.[79] "Were we able to run our nuclear plants this way they might do better, but hopefully we are not prepared to have all risky systems exist as total societies separated from normal civilian affairs."[80]

Others, however, are less optimistic about whether or not militaries are any safer from the problems that plague civilian organizations. In particular, the work of Scott Sagan and Scott Snook is important in extending Normal Accident theories to organizations, especially military ones. Together, take their insights and highlight

---

Indeed, it is for these very reasons that [NATO] regularly holds joint exercises to maintain joint operational preparedness." Roff, "The Strategic Robot Problem", p. 218. In other words, even when militaries are on the same side their systems may be prone to accidents.

[76] Gene I. Rochlin et al., "The Self-Designing High-Reliability Organization: Aircraft Carrier Flight Operations at Sea" *Naval War College Review*, Autumn 1987.

[77] Perrow, *Normal Accidents*, p. 290.

[78] Perrow, *Normal Accidents*, pp. 290-1.

[79] As noted above, this is one of the areas that scholars have suggested that there is overlap between Normal Accident Theory and High Reliability Theory. See, for example, Shrivastava, et al.

[80] Charles Perrow, "Accidents in High Risk Systems" *Technology Studies*, Vol. 1, No. 1, 1994. p. 15.

four reasons why High Reliability Theory strategies are ineffective at preventing normal accidents.

First, even when it is possible to enact the strategies that High Reliability Theorists advocate for, there are strong disincentives against exposing serious failures. For example, financial considerations and pressures from powerful elite interfere with attempts to expose problems. Sagan notes that this will influence the reporting of near-accidents by operators, the beliefs of organizational historians about the acceptable record, and the public interpretation of events by senior authorities.[81]

Second, Sagan argues that complex organizations have accidents because production pressures and parochial interests are valued over official safety goals. Put simply, getting a task done is often prized above getting a task done safely. Errors with AI may not be reported in interest of getting the technology on the battlefield. Third, the nature of military organizations make them prone to normal accidents. Sagan notes that strong organizational control over members "can encourage excessive loyalty and secrecy, distain for outside expertise, and in some cases even cover-ups of safety problems, in order to protect the reputation of the institution."[82] Additionally, within these organizations that there are conflicting interests (which leaders may only pay casual attention to when choosing between trade-offs). Further, related to the point about challenges in reporting accidents discussed above, Sagan notes there are constraints on organizational learning when there are serious disincentives against exposing existing failures. Finally, as outsiders – we should be skeptical: "One should never assume that the machine does not need to be fixed, simply because one has been told that it "ain't broke"."[83]

Finally, Snook argues that the challenge for safety is what he calls "practical drift", an organizational dynamic he describes as:

> …the slow steady uncoupling of local practice from written procedure. It is this structural tendency for subunits to drift away from globally synchronized rule-based logics of action toward locally determined task-based procedure that places complex organizations at risk.[84]

This occurs when the rules do not match the situation that individuals confront on a daily basis. "Pragmatic individuals adjust their behavior accordingly; they act in ways that better align with their perceptions of current demands. In short, they break the rules."[85] While this may seem shocking on the surface, for Snook it makes sense that this would occur in an area of military operations: "as a practical matter, whom would you listen to: the ghost of some long-forgotten planners in the form of

[81] Sagan, *The Limits of Safety*, p. 257.
[82] Sagan, *Limits of Saftey*, p. 254.
[83] Sagan, *Limits of Safety*, p. 259.
[84] Snook, *Friendly Fire*, p. 24.
[85] Snook, *Friendly Fire*, p. 193

an out dated [operational plan], or your immediate predecessor and current chief of plans?" Therefore, "as long as the system remains loosely coupled, drift continues to creep along largely unnoticed and unchecked."[86] When an incident or crisis occurs, and tight coupling re-emerges, dramatically increasing the risk for an accident as different actors follow different sets of rules.[87] In this sense, like Sagan, Snook notes that the military's strong organizational control can actually create risks, "the tighter the rules, the greater the potential for sizable practical drift to occur as the inevitable influence of local tasks take hold."[88]

Organizational culture plays an important role in both Sagan's and Snook's approaches to understanding accidents and military technology. Both are very aware of the fact that they are dealing with humans who are prone to pride and loyalty as well as error. So will this improve if we take humans out of the equation? After all, it may be argued that robots would not be subject to the pressures that Sagan and Snook identify. For example, robots could be programed to report errors to all levels of a command, and that they would not be subject to emotions such as pride.

From a Normal Accident Theory perspective, the answer is no. Even if we could make them perfect, LAWS would still be deployed from and operating in an organization impacted by the kinds of faults that Sagan and Scott recognize. Further, a more likely scenario is that autonomous systems will be operating with humans – who might be equally dismissive (or fearful) of reporting malfunctions, failures or near-misses.[89] Similarly, Sagan notes that organizations have strong disincentives against exposing serious failures that inhibits learning.[90] As such, there is every reason to believe that LAWS will be prone to normal accidents, despite strict control over their use. And while there will always be a significant degree of uncertainty surrounding all weapons, when combined with problems of attributing responsibility with LAWS, this becomes a significant problem.

The Responsible Robot

Having discussed the criticisms raised against Normal Accident Theory, and seeing how issues related to the LAWS debate may fit in, we can begin to look at some further implications of "normal accidents" for these weapon systems, especially in the area of responsibility and accountability for when things go very wrong. One of the first issues is that it is not automatically clear where responsibility should lie. Traditionally, the laws of war require that an armed force fight under a hierarchical chain of command to ensure that an individual who commits a violation may be held

[86] Snook, *Friendly Fire*, p. 195.
[87] Snook, *Friendly Fire*, pp. 199-201
[88] Snook, *Friendly Fire*, p. 201.
[89] Snook notes that "near misses" are not likely to be reported and if they are, "I am still doubtful that the right lessons would have been learned, that appropriate action would have been taken." *Friendly Fire*, p. 217.
[90] Sagan, *The Limits of Safety*, p. 257.

to account. This principle is generally known as *command responsibility*. While commanders have always been held responsible for the outcome of the orders he or she gives for centuries, following the Second World War and Additional Protocol I in 1977, a superior is now also responsible for the offences of those under his or her command if he or she knew, or ought to have known, of them and failed to take steps to prevent them.[91]

Within certain limitations, a commander is entitled to assume that orders issued by his or her superiors and the state he serves are issued in conformity with international law. To be held criminally responsible for issuing an illegal order, an intermediate commander 'must have passed the order to the chain of command and the order must be one that is criminal upon its face, or one which he is shown to have known was criminal." In other words, there must be criminal intent (*mens rea*).[92]

But is this the appropriate framework for thinking about responsibility? It might be argued that applying a command-subordinate relationship projects some responsibility onto the LAWS themselves, inadvertently anthropomorphizing them. Instead, it might be more appropriate to think about the relationship as one of direct responsibility which then places more responsibility on the human who has deployed the weapon. But this still does not get around the problem of whether a commander reasonably anticipate the actions of a LAWS because of the freedom they will have to make decisions on the battlefield. Indeed, this may make some hesitant to use LAWS in the first place.[93]

The US Department of Defense Directive seems to err on the side of the direct responsibility approach, indicating that commanders are responsible for their actions in deploying such weapons and ensuring that they are in accordance with the law of war.[94] But if these weapons have gone through an Article 36 review, it would be plausible for the individual to claim that he or she was given the weapon under the assumption that it worked and was legal, and that they had taken all feasible precautions as required by international law.[95]

Unsurprisingly, this concern has been raised by those who wish to ban LAWS: without meaningful human control it will be very difficult to hold anyone to account if an accident occurs. As Human Rights Watch argues:

> Human commanders or operators could not be assigned direct responsibility for the wrongful actions of a fully autonomous weapon,

[91] A. P. V. Rogers, *Law on the battlefield*, Second edition. Manchester: Manchester University Press, 2004. p. 189-190
[92] Rogers, *Law on the battlefield*, pp. 190-191.
[93] Grateful to XXXXXXXXX for advice on this point.
[94] DOD Directive 3000.09 "Autonomy in Weapons Systems".
[95] Feasible precautions are outlined in Article 57 of Additional Protocol I of the 1949 Geneva Conventions.

> except in rare circumstances when those people could be shown to have possessed the specific intention and capability to commit criminal acts through the misuse of fully autonomous weapons….
> A commander would nevertheless still escape liability in most cases. Command responsibility holds superiors accountable only if they knew or should have known of a subordinate's criminal act and failed to prevent or punish it.[96]

For some, this is an explicitly ethical concern. Robert Sparrow argues that if it is not possible to justly hold someone accountable for the actions of LAWS that it is unethical to use them in armed conflict. "The least we owe our enemies is allowing that their lives are of sufficient worth that someone should accept responsibility for their deaths."[97] Others focus on the potential risk of a moral hazard. As Grut notes, because LAWS challenge notions of human accountability that lie at the heart of the modern laws of war and international criminal justice, there is the possibility that the development of LAWS "might incentivize the creation of weapon systems specifically so that the state and individuals can avoid liability for the conduct of war."[98]

Scholars who argue against a ban suggest that these concerns are overwrought. Schmitt and Thurner, for example, argue that humans will always be involved with LAWS at some level. And the decision to use the weapons means that someone will always be legally accountable. "The mere fact that a human might not be in control of a particular engagement does not mean that no human is responsible for the actions of the autonomous weapon system. A human must decide how to program the system and when to launch it."[99]

From a Normal Accident Theory perspective all of these arguments are problematic. Indeed, the real issue is not that there is or is not someone in the chain of command to hold accountable, but that with accidents in tightly coupled, highly interactive system that this *responsible someone* does not exist. Accidents are the result of system characteristics, not individuals. Normal accidents:

> …represent interactions that were not in our original design of our world, and interactions that we as "operators" could not anticipate or reasonably guard against. What distinguishes these interactions is that they were not designed into the system by anybody; no one intended them to be linked. They baffle us because we acted in terms of our own

[96] Human Rights Watch, *Mind the Gap*, p. 2.
[97] Sparrow, "Killer Robots", pp. 66-67.
[98] Grut, "Challenge of Lethal Autonomous Robotics", p. 14. A similar point is raised by Horowitz who argues that there is a "real and significant risk of moral offloading," "Ethics and Morality of Robotic Warfare", *Daedalus,* forthcoming p. 12
[99] Schmitt and Thurner, "Out of the Loop", p. 270.

> designs of a world that we expected to exist – but the world was different.[100]

One of Perrow's core arguments is that that commissions of inquiries that are designed to look into accidents are often on the hunt for someone on whom blame can be placed. Indeed, "operators" are often held accountable although there may be no clear procedure to follow after an unanticipated interaction or there was no reasonable way for them to know what was going on during a crisis.[101] Further, for Perrow, efforts to point fingers at operators detracts from other culpable individuals, including system owners, or governments who regulate high-risk technology.

In his application of Normal Accident Theory to "friendly fire" incidents, Snook found that in many cases no one individual or group of persons were found responsible for several accidents that had occurred – despite months, if not years of trying. "In the case of [an accidental shootdown of a helicopter], no single cause was identified. Nothing broke and no one was to blame; yet, everything broke and everyone was to blame…"[102] Yet, this does not stop society from finding someone they can find culpable:

> "Our tendency to blame individuals for perverse outcomes of complex incidents continues to be perhaps the most consistent findings across all accident investigations I have reviewed." While acknowledging that individuals do make mistakes, in Normal Accidents these tend to be the final link in a long chain of events where removing any one link would likely have produced a very different outcome.[103]

Perrow famously describes this long chain of events in his "A Day in the Life" story. In this tale, an individual stays home from school or work because they have an important job interview downtown. However, while making breakfast, the individual finds that his or her partner has accidentally left the coffee pot on too long and the glass pot has cracked. Desperate for a coffee before the interview, the individual searches for an older coffee pot which takes longer to brew, but gets the job done. This however, has put the person in a rush and they dash out the door. Unfortunately, in their haste, he or she left their keys for the door and car in the apartment, which automatically locks. Initially the person is relieved to remember that there is a spare apartment key in the hallway, but hopes are dashed when it is remembered the key was lent out to a friend who needed to stop by last week. Increasingly desperate, the person asks a neighbour to borrow their car, but the neighbour replies that unfortunately the transmission has broken. The individual considers taking public transportation, but the neighbour notes that there is a strike

---

[100] Perrow, *Normal Accidents*, p. 75.
[101] Perrow, *Normal Accidents*, p. 53.
[102] Snook, *Friendly Fire,* p. 10.
[103] Snook, *Friendly Fire*, p. 205.

going on and there are no buses running. Using the neighbour's phone, the individual tries to call a cab, but none are available because of the bus strike. Sadly, the individual gives up trying to get downtown and cancels the job interview.

What was the cause of the failure of the individual to get to the job interview? Was it human error (leaving the coffee pot on or forgetting the keys), mechanical failure (neighbour's broken car), the environment (bus strike and taxi overload), system design (being able to lock yourself out of the apartment, lack of emergency capacity in the taxi fleet), or procedures used (warming up coffee in a glass pot, allowing only normal time in the morning to get to the interview)?

The answer for Perrow is "none of the above": "The cause of the accident is to be found in the complexity of the system. That is, each of the failures – design, equipment, operators, procedures, or environment – was trivial by itself." However, even though the failures might be trivial, they became very serious when they interacted. Therefore, "It is the *interaction* of the multiple failures that explains the accident."[104]

Bringing this back to LAWS, we can see similar issues in trying to assign responsibility. Who should be held accountable if an accident with grievous harm (such as mass civilian casualties) occurs? There seems to be consensus in the literature that holding programmers accountable is unfair and unrealistic – there would need to be evidence that they specifically intended to cause harm in programs that would likely contain millions of lines of codes.[105] The challenges of testing weapons under Article 36 weapons reviews has already been discussed. But assuming that a state put its systems through such a test, it could also argued that a state or military was acting in good faith reviewed the weapon, tested it, and even perhaps including redundant safety systems. Finally, individual units or soldiers who deployed such weapons according to the guidance given to them could equally argue that they were following the rules before a technical glitch occurred. And this does not even include confounding environmental factors.

In summary, the problem with responsibility and LAWS is not uncertainty over who to blame but the fact that there will very seldom be anyone that can be held responsible legally or justly. Tightly coupled, highly interactive systems, which are inevitably prone to Normal Accidents, diffuse responsibility across many people, over time and even across several organizations. So while some suggest "to look beyond the weapon to find the human agent responsible", this is a very problematic proposition.[106] This, in turn, poses serious challenges for how we think about restraints on war going forward – it is not about figuring out who to blame, but

---

[104] Perrow, *Normal Accidents*, p. 7.
[105] Scholars who are for and against a ban on LAWS agree on this point. See Sparrow, "Killer Robots", p. 70 and Schmitt and Thurner, "Out of the Loop", pp. 278-9.
[106] Vik Kanwar, "Post-Human Humanitarian Law: The Law of War in the Age of Robotic Warfare", *Harvard National Security Journal*, Vol. 2, 2011. pp. 617-628. p. 627

perhaps reconciling the fact that there may be no one, and dealing with the moral hazards this may raise.

Facing the Fast and Furious Future

Of course it is doubtful that a responsible military will launch a weapon if there is a high risk of failure or the consequences of using it are unknown. But, based on the findings of the discussion above, ensuring that these weapons are capable of distinguishing between legitimate military targets and civilian objects, and use force proportionately will be a considerable technical challenge. Accident investigations involving these weapons will be, at best, a problematic legal challenge. Put bluntly, we are likely moving into an era where at least part of our frameworks for governing weapons may be insufficient.

If LAWS are going to fail in ways that we cannot anticipate, and it will be nearly impossible to hold anyone to account if accidents occur, there are clearly serious implications for how we regulate these weapons. Previously, when the laws of war have been met with technical change and innovation international society has been able to adapt – either by seeing weapons through the principles that form the law or through weapons bans. For example, submarines, unmanned aerial vehicles (UAVs), and Tomahawk missiles have been able to fit into the current framework governing conventional weapons. When weapons did not fit, or seemed somehow especially abhorrent, international society dealt with them through separate treaties such as the 1925 Geneva Convention (on gas weapons), 1972 Biological Weapons Convention and 1993 Chemical Weapons Convention, and the various nuclear weapons treaties, etc.

LAWS, however, will be a challenge to regulate or ban for a number of reasons. One of the greatest challenges is that they are not likely to suddenly appear on the battlefield. There will be tactical issues to be worked out as machines and men are likely teamed up to work together on the battlefield. Figuring out how this will work in practice may take time. Most importantly, the weapons are still technically evolving – it is likely that we will see more and more platforms with autonomous parts and components in the coming years and decades. In this sense, unlike the use of gas weapons in the First World War, there will be a gradual transition over time. And it will be hard to regulate (or ban) if we do not yet know what we are dealing with.

Nevertheless, it is almost certain that this is where advance militaries are going. The US military in particular has made it clear that autonomous systems and robotics will be a part of its new "Third Offset" strategy.[107] The question is, how should

[107] See Secretary of Defence Chuck Hagel's Keynote address to the Reagan National Defense Forum, Ronald Reagan Presidential Library, Simi Valley, California, 15 November 2014. Available online: http://www.defense.gov/News/Speeches/Speech-View/Article/606635

states and international society deal with these developments? What strategies can they engage with now that will pay off later?

There has been considerable pressure from international civil society to "ban killer robots".[108] Generally, proponents of this view argue that if there are serious questions as to the legalities of the weapons (not to mention the morals and ethics of delegating the decision to kill to a machine), they should be banned like chemical, biological and other conventional weapons (landmines and cluster munitions). However, there are several practical problems with a ban that such calls will face. This is not to say that such a step would be impossible, but at least three serious challenges stand in the way of a ban being effective. First, as noted above, LAWS do not yet exist, and it will be difficult to pinpoint a time when they actually do. Determining what is and what is not a LAWS will be very difficult under these circumstances.

Second, it is one thing to tell governments not to develop autonomous systems for military applications, it is another to get private companies to do so. Indeed, in Western countries, particularly the United States, developments in artificial intelligence and robotics are coming from the private sector – they are not driven by the government. If anything, the US military has consistently had to go to Silicon Valley in order to remind companies that they too are a potential client.[109] And while private companies may not be specifically designing their inventions for military use, it will not be difficult for governments, but also private weapons manufacturers to apply such innovations to weapons systems.

Third, a ban on these systems may produce unfortunate side effects. Insisting on "meaningful human control" may push the development of technologies that some find equally unsettling, such as neuroweapons that link the minds of pilots to the machines they are controlling. Such weapons generate a similar level of legal concerns over responsibility, but would basically meet the demands by the civil society activists looking for ban on LAWS.[110] Yet, it is highly doubtful that this is their preferred outcome.

The international community is therefore left with an extremely difficult situation: nations are developing weapons that we know will fail in unanticipated ways due to

---

[108] See, for example, the bans by the Campaign to Stop Killer Robots, http://www.stopkillerrobots.org/the-solution/, Human Rights Watch https://www.hrw.org/topic/arms/killer-robots, and the International Committee for Robot Arms Control http://icrac.net/call/

[109] Brian Fung, "The huge issue that's keeping Silicon Valley and the Pentagon apart" *Washington Post*, 10 June 2016. Available online: https://www.washingtonpost.com/news/the-switch/wp/2016/06/10/the-pentagon-wants-to-cozy-up-to-silicon-valley-heres-one-big-thing-keeping-them-apart/. See also the discussion on difficulties with procurement in Singer, P. W. Singer, *Wired For War: The Robotics Revolution and Conflict in the 21st Century*, New York: Penguin, 2009. pp. 254-260.

[110] White "Brave New World". pp. 177-210.

their unprecedented complexity, and where prosecution for accidents or even war crimes will be difficult. At this stage developing a strict legal framework is problematic or may actually cause more problems. And yet, the idea that the international community does nothing is equally if not more unappealing. So what is to be done?

Understanding LAWS and their implications through a Normal Accident Theory framework can help guide some preliminary steps that both governments and civil society can take. First, efforts need to be made in developing norms and laws regarding responsibility in systems where it is inherently diffused across a number of actors. This will help ameliorate the moral hazard created if states are able to design and use weapons where accountability is basically impossible. It is not realistic to put the entire military of a country on trial. At the same time, there is a problem of fairness if no one can ever be held to account reasonably. So what could such a system look like? Over the last 60 years in situations of organised mass killing international courts have had to determine the individual responsibility for many actors ranging from prison guards and bureaucrats to political decision-makers. The situation with LAWS will almost certainly be different from situations of mass-killing, but depending on the severity of an incident, such trials present models for thinking about responsibility across a number of actors.

Second, although we should expect "normal accidents" with LAWS, this should not stop us from developing norms and standards to ensure that weapons fail safely. While Normal Accident Theorists are very pessimistic (and many might advocate a ban on LAWS) none would advocate shirking safety and procedure. In this way, Normal Accident Theory is a reminder of the limits of redundancy, safety devices and organizations – and that there will be limits to our ability to predict problems and what might eventually go wrong.

Third, private industry must be brought into discussions over the regulations and banning of such weapons. As Peter Singer notes in his book on military robotics, not one robotics research, developer, program manager referenced the laws of war in the hundreds of interviews he conducted. "That is, not a single organization, research lab, or company working on robotics today is formally linked up with the [International Committee of the Red Cross] or has in place… reviews… necessary for new weapons."[111] While some efforts are being made in this regard by the International Committee for Robot Arms Control, more effort to bring the private sector into discussions that are typically between humanitarian organizations, international civil society and governments will be needed. Getting innovators, humanitarians and governments talking at the same table is extremely important.

Finally, there must be continued diplomacy and discussion as technological developments arise. One of the most important things that states can do its to

[111] Singer, *Wired for War*, p. 385. See also the discussion in Grut, "Challenge of Autonomous Lethal Robotics", pp. 20-21.

present and discuss their policies and regulations as well as their concerns. This will be a challenge for the national security apparatuses of many states, who typically wish to discuss nothing "for fear of revealing capabilities or programming details to adversaries, or inviting industrial espionage and reverse engineering of systems." And yet Anderson and Waxman note that despite the risks there are larger issues at stake, including shaping the normative terrain and countering regulation that would be unrealistic and ineffective, or too generous.[112] Much of this dialogue is taking place in forums such as the annual meeting of the Convention on Conventional Weapons (CCW) in Geneva. However, discussions should also be taking place at other forums, such as international scientific gatherings.

In the end international society needs to brace itself for the fact that LAWS will not fit well under our current framework for governing weapons. And unfortunately, doing something about this, for now, will be extremely hard. In the meantime, Normal Accident Theory provides a useful framework through which to understand the challenges the world is likely to face, including the fact that these weapons are certain to fail in ways that we cannot anticipate and that accountability will be problematic at best. A better understanding of these issues enables international society to begin to take steps, outlined above, in addressing the quandaries that LAWS will present.

[112] Kenneth Anderson and Matthew Waxman, "Brave New War", Defining Ideas, 14 December 2012. Available online: http://www.hoover.org/research/brave-new-war . See also, Anderson and Waxman, "Law and Ethics for Robot Soldiers", p. 16 and "Law and Ethics for Autonomous Weapon Systems", p. 23.

Bibliography

Amodei, Dario et al. "Concrete Problems in AI Safety", Available online: https://arxiv.org/abs/1606.06565

Anderson, Kenneth and Matthew Waxman. "Brave New War", Defining Ideas, 14 December 2012. Available online: http://www.hoover.org/research/brave-new-war

Anderson, Kenneth and Matthew Waxman. "Law and Ethics for Autonomous Weapons Systems: Why a Ban Won't Work and How the Law of War Can", Hoover Institution, 9 April 2013. Available online: http://www.hoover.org/research/law-and-ethics-autonomous-weapon-systems-why-ban-wont-work-and-how-laws-war-can

Anderson, Kenneth and Matthew Waxman. "Law and Ethics for Robot Soldiers", *Policy Review*, 1 December 2012. Available online: http://www.cfr.org/world/law-ethics-robot-soldiers/p29598

Asaro, Peter. "On banning autonomous weapon systems: human rights, automation and the dehumanization of lethal decision-making", *International Review of the Red* Cross, Vol. 94, No. 886, Summer 2012. pp. 687- 709

Bradshaw, Jeffrey M., Robert R. Hoffman, Matthew Johnson and David D. Woods. "The Seven Deadly Myths of "Autonomous Systems"", *IEEE Intelligent Systems*, May/June 2013. pp. 2-9.

Bekey, George A. *Autonomous Robots: From Biological Inspiration to Implementation and Control*, Cambridge, MA: The MIT Press, 2005.

Campaign to Stop Killer Robots https://www.stopkillerrobots.org/

Clooney, Michael. "Beyond the petaflop: DARPA wants quintillion-speed computers", *Network World*, 23 June 2010. Available online: http://www.networkworld.com/article/2231123/security/beyond-the-petaflop--darpa-wants-quintillion-speed-computers.html

Department of Defense Directive Number 3000.09, "Autonomy in Weapons Systems", 21 November 2012. Available online: http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf

Executive Office of the President National Science and Technology Council Committee on Technology. "Preparing for the Future of Artificial Intelligence", October 2016. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Fung, Brian. "The huge issue that's keeping Silicon Valley and the Pentagon apart" *Washington Post*, 10 June 2016. Available online:

https://www.washingtonpost.com/news/the-switch/wp/2016/06/10/the-pentagon-wants-to-cozy-up-to-silicon-valley-heres-one-big-thing-keeping-them-apart/

Gershgorn, David. "Fooling the Machine", *Popular Science*, 30 March 2016. Available online: http://www.popsci.com/byzantine-science-deceiving-artificial-intelligence

Grut, Chantal. "The Challenge of Autonomous Lethal Robotics to International Humanitarian Law", *Journal of Conflict and Security Law,* Vol. 18, No. 1, 2013. pp. 5-23

Hagel, Chuck. "*Keynote address to the Reagan National Defense Forum*," Ronald Reagan Presidential Library, Simi Valley, California, 15 November 2014. Available online: http://www.defense.gov/News/Speeches/Speech-View/Article/606635

Hopkins, Andrew. "Was Three Mile Island a 'Normal Accident'?" *Journal of Contingencies and Crisis Management*, Vol. 9, No. 2, June 2001. pp. 65-72.

Horowitz, Michael C. and Paul Scharre, "The Morality of Robotic War", *New York Times*, 27 May 2016. Available online: http://www.nytimes.com/2015/05/27/opinion/the-morality-of-robotic-war.html

Horowitz, Michael C. "The Ethics and Morality of Robotic Warfare: Assessing the Debate Over Autonomous Weapons", *Daedalus*, *FORTHCOMING.*

Human Rights Watch, *Mind the Gap: The Lack of Accountability for Killer Robots*, 2015. Available online: https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots

Human Rights Watch https://www.hrw.org/topic/arms/killer-robots

International Committee for Robot Arms Control http://icrac.net/

Jenks, Chris. "False Rubicons, Moral Panic & Conceptual Cul-De-Sacs: Critiquing, Reframing the Call to Ban Lethal Autonomous Weapons", 2016. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2736407##

Kalra, Nidhi and Susan M. Paddock. *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?*, Washington DC: RAND*,* 16 April 2016. Available online: http://www.rand.org/pubs/research_reports/RR1478.html

Kanwar, Vik. "Post-Human Humanitarian Law: The Law of War in the Age of Robotic Warfare", *Harvard National Security Journal*, Vol. 2, 2011. pp. 617-628.

La Porte, Todd R. "A strawman speaks up: Comments on *The Limits of Safety*", *Journal of Contingencies and Crisis Management*, Vol. 2, No. 4, 1994. pp. 207-211

La Porte, Todd R. and Paula M. Consolini. "Working in practice but not in theory: theoretical challenges of "high reliability organizations", *Journal of Public Administration Research and Theory*, Vol 1. No. 1, 1991. pp. 19-47.

Lamport, Leslie, Robert Shostak and Marshall Pease. "The Byzantine Generals Problem", ACM Transactions on Programming Languages and Systems, Vol. 4, No. 3, July 1982. pp. 382-401.

Le Coze, Jean-Christophe. "1984-2014. Normal Accidents. Was Charles Perrow Right for the Wrong Reasons?", *Journal of Contingencies and Crisis Management*, Vol. 23, No. 4, December 2015. pp. 275-286. p. 284. See also Sagan, *The Limits of Safety*; Snook, *Friendly Fire*.

Marchant, Gary E. et al. "International Governance of Autonomous Military Robots", *The Columbia Science and Technology Law Review*, Vol. XII, 2011. pp. 272-315

Metz, Cade. "In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go", *Wired*, 27 January 2016. Available online: https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/

Morkevicius, Valerie. "Tin Men: Ethics, Cybernetics and the Importance of Soul", *Journal of Military Ethics*, Vol. 13, No. 1, 2014. pp. 3-19

Perrow, Charles. "Accidents in High Risk Systems" *Technology Studies*, Vol. 1, No. 1, 1994.

Perrow, Charles. "A Personal Note on *Normal Accidents*", *Organization & Environment*, Vol. 17, No. 1, March 2004. pp. 9-14.

Perrow, Charles. "Difficulties with Network Centric Warfare" in Jacques S. Gansler, Hans Binnendijk, *Information Assurance: Trends in Vulnerabilities, Threats and Technologies*, Washington DC: National Defense University, 2004. pp. 139-146

Perrow, Charles. *Normal Accidents: Living with High-Risk Technologies*, Princeton: Princeton University Press, 1999.

Perrow, Charles. "The Limits of Safety: The Enhancement of a Theory of Accidents", *Journal of Contingencies and Crisis Management*, Vol. 2, No. 4, 1994. p. 212-220.

Rijpma, Jos. A. "From Deadlock to Dead End: The Normal-Accidents-High Reliability Debate Revisited", *Journal of Contingencies and Crisis Management*, Vol. 11, No. 1, March 2003. pp. 37-45.

Rochlin, Gene I. et al. "The Self-Designing High-Reliability Organization: Aircraft Carrier Flight Operations at Sea" *Naval War College Review*, Autumn 1987.

Roff, Heather M. and Richard Moyes. "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons", Article 36, April 2016. Available online: http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf

Roff, Heather M. "The Strategic Robot Problem: Lethal Autonomous Weapons in War", *Journal of Military Ethics*, Vol. 13, No. 3, 2014. pp. 211-227

Rogers, A. P. V. *Law on the battlefield*, Second edition. Manchester: Manchester University Press, 2004.

Rosa. Eugene A. "Celebrating a Citation Classic – and More", *Organization & Environment*, Vol. 18, No. 2, June 2005. Pp. 229-234.

Sagan, Scott D. *The Limits of Safety: Organizations, Accidents and Nuclear Weapons*, Princeton: Princeton University Press, 1993.

Scharre, Paul and Michael C. Horowitz, "An Introduction to Autonomy in Weapon Systems" Centre for New American Security Working Paper, February 2015. Available online: http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf

Scharre, Paul. "Autonomous Weapons and Operational Risk" Washington D.C.: Centre for New American Security, February 2016. Available online: http://www.cnas.org/sites/default/files/publications-pdf/CNAS_Autonomous-weapons-operational-risk.pdf.

Scharre, Paul. "Between a Roomba and a Terminator: What is Autonomy?", *War on the Rocks,* 18 February 2015. Available online: http://warontherocks.com/2015/02/between-a-roomba-and-a-terminator-what-is-autonomy/

Scharre, Paul. "Why Unmanned?", *Joint Forces Quarterly*, No. 61, 2nd Quarter, 2011. pp. 89-93

Schmitt, Michael N. and Jeffrey S. Thurnher, ""Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict", *Harvard National Security Journal*, Vol. 4, 2012-2013. pp. 231-281

Shanahan, Murray. *The Technological Singularity*, Cambridge, MA: MIT Press, 2015. pp. 55-58

Sharkey, Noel. "Saying 'No!' to Lethal Autonomous Targeting", *Journal of Military Ethics*, Vol. 9, No. 4, 2010. pp. 369-383

Sharkey, Noel. "The evitability of autonomous robot warfare", *International Review of the Red Cross*, Vol. 94, No. 886, Summer 2012

Shrivastava, Samir, Karan Sonpar, and Federica Pazzaglia.Normal Accident Theory versus High Reliability Theory: A resolution and call for an open systems view of accidents", *Human Relations*, Vol. 62. No. 9, 2009. pp. 1357-1390.

Silvast, Antti and Ilan Kelman. "Is the Normal Accidents perspective falsifiable?", *Disaster Prevention and Management,* Vol. 22, No. 1, 2013. pp. 7-16.

Simons, Ned. "Britain Not Building 'Killer Robots', Minister Insists", *HuffingtonPost UK*, 18 June 2013. Available online: http://www.huffingtonpost.co.uk/2013/06/18/killer-robots-britain_n_3459463.html

Singer, P. W. *Wired For War: The Robotics Revolution and Conflict in the 21st Century*, New York: Penguin, 2009.

Snook, Scott A. *Friendly Fire: The Accidental Shootdown of U.S. Blackhawks over Northern Iraq*, Princeton: Princeton University Press, 2000.

Sparrow, Robert. "Killer Robots", *Journal of Applied Philosophy*, Vol. 24, no. 1, 2007. pp. 62-77

Stockholm International Peace Research Initiative, "Implementing Article 36 weapon reviews in the light of increasing autonomy in weapon systems", 11 November 2015. Available online: https://www.sipri.org/media/press-release/2015/implementing-article-36-weapon-revews-light-increasing-autonomy-weapon-systems.

Thomsen, Michael. "Microsoft's Deep Learning Project Outperforms Humans In Image Recognition", Forbes, 19 February 2015. https://www.forbes.com/sites/michaelthomsen/2015/02/19/microsofts-deep-learning-project-outperforms-humans-in-image-recognition/#49ef8484740b

Turner, Brian A. and Nick Pidgeon, *Man-Made Disasters* (second ed.), Oxford: Butterworth-Heinemann, 1997

Vaughn, Diane. *The Challenger launch Decision: Risky Technology, Culture and Deviance at NASA*, Chicago: University of Chicago Press, 1996.

White, Stephen E. "Brave New World: Neurowarfare and the Limits of International Humantarian Law", *Cornell International Law Journal*, Vol. 41, Issue 1, Winter 2008. pp. 177-210.