Aaron Gordon
CST 300 Writing Lab
October 18, 2016

Fight or Flight: On the Ethics of an Autonomous Weapons Ban

Augustine of Hippo, known more commonly as St. Augustine, was a prolific theologian and philosopher of late antiquity and one of the earliest writers on the ethics of war. Augustine articulated the conditions under which a nation-state may be justified in waging war, namely as a response to injustice and in pursuit of peace. However, as Langan (1984) notes, Augustine's theory of just war "does not include non-combatant immunity" (p. 19).

Nearly a millennium-and-a-half after Augustine lived, the articles of the *Convention for the Amelioration of the Condition of the Wounded in Armies in the Field* (1864) - popularly named "The Geneva Convention" after its host-city - asserted the wartime immunity of non-combatant medical personnel. The third and fourth Geneva Conventions would make provisions for the protection of prisoners of war and civilians, respectively, during wartime (Rules and Conventions, n.d.). The Geneva Conventions would call for many of the stipulations present in modern international law, including perhaps most famously a ban on gas and other biological weapons.

By the second half of the 20th century the doctrine of proportionality was propelled by the horrors of the second world war into the international discourse on human rights (Taskovska, 2012). The doctrine of proportionality was concerned with preventing the return of totalitarian regimes to Western Europe; it began as a claim of policy arguing governments create laws which stipulate *reasonable* punishments for their violation. Shoplifting, to give a naive example, should result in a fine, not the death sentence. Proportionality would quickly be adopted into

discussions on international human rights and the ethics of war. Bryen (2014) succinctly summarizes the role of proportionality in war, writing "Proportionality weighs the necessity of a military action against suffering that the action might cause to enemy civilians in the vicinity."

As technology marches forward, the importance of proportionality becomes evident. The indiscriminate killing power of gas weapons used during the First World War led to their aforementioned ban by the Geneva Conventions. The civilian casualties caused by fields of anti-personnel landmines - explosives designed to detonate when impressed with the weight of an average human - which have remained active long after the conflicts that saw their deployment, prompted a United Nations ban in 1997 (Landmines, n.d.). More recently, heat-seeking missiles have been developed which "hunt down" their target after being launched.

Each of the above weapons may be deployed by a military without a specific target in mind: gas blankets an entire area, mines wait for soldiers to come to them and heat-seeking missiles find their own targets. Now a new generation of weapons are leveraging the powers of computer vision and artificial intelligence to identify, prioritize and destroy targets entirely on their own. These new weapons have been dubbed "autonomous" and they range from artillery cannons which automatically identify and fire upon enemy missiles to aerial vehicles such as the Salty Dog 502 which "launch, land and refuel in midair without human intervention" (Cohen, 2013). Other autonomous weapons called loitering munitions are, as aptly described by Horrowitz and Scharre (2015, February), "launched into a general area where they will loiter, flying a search pattern looking for targets within a general class, such as enemy radars, ships or tanks." (p. 19). When a target of the programmed class is found, the loitering munition attacks.

Autonomous weapons are described as leaving humans "out of the loop" because targeting and firing decisions are made entirely apart from a human operator. Much like gas and landmines, autonomous weapons are deployed then left to their own ends. Also much like gas and landmines, the ethical implications of autonomous weapons must be carefully considered. Already multiple groups have voiced opinions on the topic of autonomous weapons; opinions which generally fall into one of three categories: ban all autonomous weapons, require humans "in the loop" or delay legislation.

The argument for an outright ban on autonomous weapons is best exemplified by an open letter from the Future of Life Institute. The Future of Life Institute is a self-described "outreach organization" that believes technology has become so powerful, and the consequences of its misapplication so severe, that humanity can no longer risk learning to rightly wield it via trial and error. As an organization founded and run primarily by academicians, the Future of Life Institute values free, open societies that foster learning, and works to secure the future of such societies from the harm that may be inflicted by the misuse of new technologies, such as autonomous weapons.

The open letter, titled *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, was presented in 2015. In it, the Future of Life Institute articulates the argument that autonomous weapons will become cheap weapons of mass destruction abused by tyrants - the "Kalashnikovs of tomorrow" (Future of Life Institute, 2015). Autonomous weapons, states the letter, will eventually cost so little to produce they will be readily accessible to terrorists and dictators to use in perpetrating human rights crimes. The Institute for Life thus concludes their letter with a call for an international ban on autonomous weapons, implicitly claiming that such a

ban is both enforceable and an effective means of preventing the proliferation of dangerous autonomous weapons. The open letter has a long list of signatories, including such household names as Stephen Hawking and Elon Musk.

The argument of the Future of Life Institute's letter rests on a consequentialist framework of ethics. Consequentialist ethics is focused primarily on the effect of a moral decision, in this case the effect of banning or not banning autonomous weapons. It is not surprising that engineers and physicists would construct a letter preoccupied with a cause and effect analysis of the situation. Indeed, the letter performs the sort of "moral calculus" typical of utilitarianism - a school of consequentialist ethics that asserts the morally right decision is that which minimizes pain and maximizes pleasure in a population. The Future of Life Institute reasons that, clearly, banning autonomous weapons avoids a great deal of pain for a great many people while failure to ban offers little promise of pleasure to anyone but a handful of warlords.

The United Nations hosts an annual Convention on Certain Conventional Weapons, or CCW, which consists, in part, of the Meeting of Experts on Lethal Autonomous Weapons Systems, or LAWS. As a part of the United Nations, the CCW values the upholding of international human rights laws. The CCW has inherited the prestigious legacy of the Geneva Conventions, and seeks to continue the tradition of pursuing the safety of civilians and soldiers through the restriction of warfare. The United Nations makes familiarity with human rights violations their business, and through the Meeting of Experts on LAWS the CCW hopes to prevent autonomous weapons from implication in any such future violations.

The Meeting of Experts on LAWS is an open forum of discussion with many, sometimes opposing, arguments made. There is a consensus among attendees, however, that human out of

the loop autonomous weapons may not conform to the doctrine of proportionality. Lt. Col. Ford (2016) broached the subject at the 2016 CCW when he stated:

> [Proportionality] is operationalized in Article 57 which requires "constant care" be taken to "spare the civilian population, civilians and civilian objects." The Protocol does not define "constant care," but this phrase suggests something more than a one-time obligation.

Ford's implication is that the care taken when programming an autonomous weapon does not satisfy the obligation to *constant* care; programming a weapon responsibly is "one-time" in nature. A fellow attendee of Ford's to the 2016 CCW, Leiblich (2016), renames this obligation "continuous discretion" and argues that international human rights law requires military personnel "exercise discretion up to the last moment before pulling the trigger". Leiblich and Ford represent the viewpoint common amongst attendees of the CCW that autonomous weapons, capable of firing without human intervention, violate the mandate that militaries practice "constant care", or "continuous discretion". The common conclusion, then, is that autonomous weapons must require a human in (rather than out of) the loop - an operator with the final say in any targeting decision made by the weapon - as humans alone are capable of exercising the constant care required by the law.

The argument presented by both Ford and Leiblich, and common to the CCW, relies upon an absolutist framework of ethics. Absolutism deems a moral decision right or wrong based not upon the effects of the decision but by whether or not the decision holds to a code of higher moral authority. In this instance the moral authority is international human law. This is a highly appropriate ethical framework for an international body of policymakers, such as the United

Nations. The CCW attendees argue that the use of human out of the loop autonomous weapons, regardless of impact, is not in alignment with the legal duties, and by deontological extension moral obligations, a military has to international human law, rendering it thus unethical. There is as well a subtle hint of Kantian thought in the argument for continuous discretion. Kant was an absolutist who posited that moral authority rested not in law, natural or divine, but rather within man himself. Leiblich's argument that an autonomous weapon cannot exercise continuous discretion is perhaps born from a Kantian notion that moral authority is uniquely human.

While the Future of Life Institute argues for an outright ban on autonomous weapons and the United Nations deliberates the ability of autonomous weapons to properly exercise constant care in war, there remains a large contingent of professionals and academics opposed to legislation of any form at the present time. This group values the safety of both military personnel and civilians in war zones and sees potential in autonomous weapons to reduce casualties on all sides of a military conflict. For these individuals, premature legislation closes a door to an unexplored, promising new field.

The arguments against legislation are as varied as the individuals who make them, but common themes do occur. Scharre and Horowitz (2015, August) raise the oft-heard argument that many of the terms appearing in the literature on the ethics of autonomous weapons, terms such as "meaningful human control" and even "autonomous weapon" itself, remain poorly defined. There exist degrees of autonomy, argue Scharre and Horowitz, which many currently deployed weapons have implemented without incident, and which must be taken into account when discussing a ban. Arkin (2015) represents the popular viewpoint that through advances in artificial intelligence autonomous weapons may someday soon prove better stewards of

international human rights law than emotional and vengeful soldiers are, and that a ban now precludes such potentially fruitful research. Yet perhaps the most frequently shared argument against a ban, echoed by Scharre, Horowitz, Arkin and throngs of commentators and commenters alike, is that it is simply unenforceable. Unlike the rare materials required to build nuclear or chemical weapons, goes the line of reasoning, the ubiquity of the components used to fabricate an autonomous weapon make it effectively impossible to regulate their construction.

These arguments are made by distinct persons, rather than a larger organization, and while their premises and claims vary, they build upon a common foundation of virtue ethics. Virtue ethics is the belief that if a man is virtuous - courageous and wise and compassionate - he will know the morally right from the morally wrong, therefore men wishing to lead moral lives are most successful when seeking virtuous lives. It is an ethical framework rooted in individualism, and so its presence in arguments formed by individuals should come as no surprise. Arguments against a ban on autonomous weapons claim that if man's goals are virtuous - the preservation of both combatant and civilian lives - then man will know the morally right way to explore this fledgling technology; the moral absolutism of international law is not needed.

Such an optimistic view of human nature is fully rejected by the argument for a complete ban on autonomous weapons, which ignores any potential benefits autonomous weapons may yield and assumes a pessimistic view of man's ability to self-regulate. As such, the argument inevitably reduces to little more than a slippery slope fallacy: if autonomous weapons are not banned, a bleak future of genocide and ethnic cleansing by robots is inevitable. It's an argument that relies far too heavily on pathos and sounds more akin to a science fiction film than reasoned

rhetoric. The Future of Life Institute's open letter attempts to prop up this invalid argument with an impressive list of signatories, but this is only argumentum ad populum - the sheer volume of signatories makes no difference in the validity of the argument - and a dubious appeal to authority - Hawking and Musk, while no doubt intelligent, are neither experts on human rights law nor even particularly qualified as ethicists. Moreover, provisions for the practical enforcement of a ban remain missing. Unlike the mutually assured destruction promised by nuclear weapons or the proven, horrific civilian casualties caused by anti-personnel mines, nations have little motivation to follow a ban on autonomous weapons, especially when the only atrocities yet committed with them remain fictional. The argument for an outright ban, particularly as presented by the Future of Life Institute, is too logically flawed to consider.

The CCW's proposal to require human intervention in all autonomous weapons systems is a much more compelling argument, though not without weaknesses. The argument fails to address the practicality of enforcement, though given the deontological nature of the CCW's discussions this is perhaps forgivable. Less easily overlooked is the assumption that autonomous weapons are incapable of exercising the sort of "constant care" required by international law. Here the CCW commits the fallacy of "begging the question" insofar as their arguments have failed to address the important and controversial assumption that machines cannot be adequately programmed to uphold human rights. Indeed, it seems much more plausible to program a machine to risk its own destruction by erring towards the protection of human rights than to ask a human to do the same. Failure to prove this assumption leaves the argument valid, but unsound. The CCW also fails to address the issue of human out of the loop autonomous weapons which

serve purely defensive roles, perhaps reinforcing the argument that the vernacular of the field is as yet too poorly defined to rightly legislate.

Considering these shortcomings in the arguments for legislation, and in view of the potential benefits autonomous weapons may yet yield, it is this author's position that, at this present time, no legislation should be adopted because no legislation is ethically required.

One must ask, what is the goal of war? To inflict maximum property destruction, maximum loss of life, upon the enemy? Certainly not! Rather, the goal of war is to prevent some greater evil than the war itself will inflict, thus the cruciality of the doctrine of proportionality. Today the United States military enforces proportionality by requiring a great degree of certainty and legal authorization from off-site JAG officers before firing on potential enemies, going so far as to pursue a "Zero Civilian Casualties" policy in its engagements in the middle east (Wong, 2015). Maintaining proportionality in the asymmetric warfare found on many current, civilian-heavy battlegrounds calls for precise, surgical strikes - not the brash carpet bombings of prior wars - and restraint in the heat of battle. Banned weapons, such as gas and landmines, do not wait for JAG authorization and do not discriminate between civilian and combatant - and this is precisely why they are banned.

Autonomous weapons, however, are very capable of meeting the requirements for proportionality. While anti-personnel mines are banned, anti-vehicle mines are not, for the simple reason that they require a much heavier weight to detonate and so, ostensibly, discriminate between individuals (which may be civilians) and much heavier vehicles (presumably tanks, though too often civilians fleeing war zones in cars). Autonomous weapons are capable of identifying and analyzing threats with much greater subtlety and on many more

dimensions than weight, which leads one to conclude they should be banned no more readily than anti-vehicle mines. Moreover, autonomous weapons are capable of accurate targeting, and do not blanket an area with death the way banned chemical and biological weapons are capable of. Perhaps most importantly, though, an autonomous weapon is patient - it will never become "trigger happy" in a tense situation. Unlike any weapon before it, the autonomous weapon, armed with the capability of thought, however rudimentary, has the potential to refuse to kill. Autonomous weapons are so readily capable of meeting the requirements of proportionality that there is no genuine ethical requirement for the passing of a ban.

To the contrary, autonomous weapons are not only able to adhere to proportionality, they may yet make the battlefield a safer environment for all sides involved. A programmatic JAG routine onboard an autonomous weapon could intelligently authorize or deny strikes in real time and would have access to far more information than current, off-site JAG officers have via radio. Autonomous weapons could be programmed to lead military or even law enforcement personnel into potentially dangerous situations. Vigilant loitering munitions could wait tirelessly for enemy military actors to expose themselves outside of civilian populations, or actively seek and destroy enemy anti-aircraft weapons to create safe zones for cargo planes delivering relief to civilians. These examples are only a few of dozens proposed in the technology's infancy; no doubt countless more will come if the technology is permitted to mature.

It is granted that the argument against legislation at this point makes certain assumptions. First, the assumption that warfare will continue to evolve away from the classical model of two armies on a field and towards the asymmetric battlefields seen today. This seems like a reasonable assumption at this time given the current political climate of the world, but could

quickly change with the advent of a war between superpowers. Second, the assumption that autonomous weapons can be adequately secured against "cyber attacks", which if proved unequivocally untrue must certainly impact the conclusions drawn.

This author readily admit to biases which may impact his conclusions. He is a citizen of a first world country which would have the greatest access to autonomous weapons once developed. Simultaneously, as a citizen of the first world, he has witnessed the cultural fallout that can occur when law enforcement make human errors in threat detection. Perhaps most importantly, this author is a computer science major that naturally wonders if computers truly cannot become more accurate at threat detection and more diligent in the preservation of human rights than we ourselves are.

There are, of course, limitations to the argument against legislation. Waiting to legislate may open the door for an autonomous weapons arms race between major world powers - although it should be noted that such arms races have occurred before without escalation to actual conflict. Moreover, not legislating is not a final solution. As the field of autonomous weapons continues to develop and the implications of the technology become better understood, a need for legislation of some kind will inevitably be necessary (i.e. can civilians own autonomous weapons?). What such legislation will look like, however, simply cannot be known in any capacity at this time. With this in mind, the call to belay legislation stands.

It is highly plausible that the United Nation's CCW's proposal for requiring humans in the loop of autonomous weapons will prove the proper legislative course. The arguments presented at the CCW are strong, and rely on a Kantian ideal of humanity that should not be undervalued. Requiring a human in the loop may prevent misfires from "bugs" in threat

detection, assuming human operators are well-trained enough not to overly rely on the weapon's recommendations.

When Augustine put ink to page 1600 years ago, he could never have dreamed up a world where the weapons themselves could reason. Autonomous weapons present a great deal of promise for the future of human rights in wartime. It is true that such weapons also present a great deal of danger, and often the temptation to attempt to rid the world of anything dangerous can be great. To legislate now, however, would require knowledge of a technology still being developed. Should our predictions of the necessary legislation prove wrong, all we will have accomplished is to have denied ourselves the opportunity to find new ways of preserving human life amidst the horror of war. Indeed, if autonomous weapons have genuine potential to save human lives then it would seem their continued development is not only morally permissible, but a moral imperative.

References

Arkin, R.C. (2015, August 5). Warfighting robots could reduce civilian casualties, so calling for a ban now is premature. *IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/autonomous-robotic-weapons-could-reduce-civilian-casualties

Bryen, S. (2014, July 20). The doctrine of proportionality. *Gatestone Institute International Policy Council*. Retrieved from https://www.gatestoneinstitute.org/4462/proportionality-doctrine

Cohen, D.F. (2013, July 25). Drones off the leash. *U.S. News & World Report*. Retrieved from http://www.usnews.com/opinion/articles/2013/07/25/autonomous-drones-and-the-ethics-of-future-warfare

Convention for the amelioration of the condition of the wounded in armies in the field. (1864, August 22). Retrieved from https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Treaty.xsp?action=openDocument&documentId=477CEA122D7B7B3DC12563CD002D6603

Ford, C. M. (2016, April). *CCW Remarks*. Paper presented at the 2016 meeting of experts on LAWS, Geneva, Switzerland. Retrieved from http://www.unog.ch/80256EDD006B8954/(httpAssets)/D4FCD1D20DB21431C1257F9B0050B318/$file/2016_LAWS+MX_presentations_challengestoIHL_fordnotes.pdf

Future of Life Institute. (2015, July 28). *Autonomous weapons: an open letter from AI & robotics researchers*. Retrieved from http://futureoflife.org/open-letter-autonomous-weapons/

Horowitz, M. C., & Scharre, P. (2015, February). *An introduction to autonomy in weapon systems*. Retrieved from https://s3.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf

Horowitz, M. C., & Scharre, P. (2015, August). Ban or no ban, hard questions remain on autonomous weapons. *IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/automaton/robotics/military-robots/ban-or-no-ban-hard-questions-remain-on-autonomous-weapons

Landmines. (n.d.). Retrieved from https://www.un.org/disarmament/convarms/landmines/

Langan, J. (1984). The elements of St. Augustine's just war theory. *The Journal of Religious Ethics, 12 (1)*. Retrieved from http://www.jstor.org/stable/40014967

Lieblich, E. (2016, April). *Autonomous weapons systems and the obligation to exercise discretion*. Paper presented at the 2016 meeting of experts on LAWS, Geneva, Switzerland. Retrieved from http://www.unog.ch/80256EDD006B8954/(httpAssets)/4FFC1EF9102E0017C1257F9A00465A01/$file/2016_LAWS+MX+Presentations_HRandEthicalIssues_Eliav+Lieblich+note.pdf

Phalanx close-in weapon system. (n.d.). Retrieved from http://www.raytheon.com/capabilities/products/phalanx/

Rules and conventions. (2014). Retrieved October 9, 2016, from http://www.bbc.co.uk/ethics/war/overview/rules.shtml

Taskovska, D. (2012). *On historical and theoretical origins of the proportionality principle*. Retrieved from http://law-review.mk/pdf/04/Dobrinka%20Taskovska.pdf

Wong, K. (2015, June 24). US aim for 'zero civilian casualties' draws criticism. *The Hill*. Retrieved from http://thehill.com/policy/defense/policy-strategy/245932-us-aims-for-zero-civilian-casualties-in-war-vs-isis