

The Future of War: Could Lethal Autonomous Weapons Make Conflict More Ethical?

Steven Umbrello, Institute for Ethics and Emerging Technologies

Phil Torres, Project for Future Human Flourishing

Angelo F. De Bellis, University of Edinburgh

Abstract: Lethal Autonomous Weapons (LAWs) are robotic weapons systems, primarily of value to the military, that could engage in offensive or defensive actions without human intervention. This paper assesses and engages the current arguments for and against the use of LAWs through the lens of achieving more ethical warfare. Specific interest is given particularly to ethical LAWs, which are artificially intelligent weapons systems that make decisions within the bounds of their ethics-based code. To ensure that a wide, but not exhaustive, survey of the implications of employing such ethical devices to replace humans in warfare is taken into account, this paper will engage on matters related to current scholarship on the rejection or acceptance of LAWs—including contemporary technological shortcomings of LAWs to differentiate between targets and the behavioral and psychological volatility of humans—and current and proposed regulatory infrastructures for developing and using such devices. After careful consideration of these factors, this paper will conclude that only ethical LAWs should be used to replace human involvement in war, and, by extension of their consistent abilities, should remove humans from war until a more formidable discovery is made in conducting ethical warfare.

Keywords: lethal autonomous weapons; artificial intelligence; military robots; ethics; laws of war

1. Introduction

The present paper evaluates the use of lethal autonomous weapons (LAWs). It argues that once LAWs reach a certain level of technical sophistication—specified below—they could become the *only* ethical means for waging war. To be clear about this point, we do not advocate for the use of violence to resolve interstate conflicts, nor do we believe that LAWs are, or could be, a definitive solution to the conundrums of “ethical warfare.” However, we do accept the following two claims: first, despite the historical decline of many forms of violence, including interstate and civil wars,ⁱ armed conflicts will continue to be fought in the coming decades and centuries. Second, governments are not going to stop developing robotic devices such as LAWs for military purposes. There are currently no international treaties that prohibit or regulate the development and deployment of LAWs, and numerous countries—most notably the United States, China, Russia, and the UK (Kirk 2015)—are actively pursuing novel LAW technologies. It follows that there is a nontrivial, and perhaps a significant, probability that states will employ LAWs in future conflicts, and this makes understanding the various ethical and legal issues surrounding their use both important and urgent.

There are two general types of LAWs: offensive and defensive. The most advanced weapons of this sort today are equipped with autonomous systems that enable them to both defend themselves and identify potential targets without human intervention, although the final decision to use lethal force remains within human control (Heyns 2013, §45). For this reason, contemporary LAWs are not yet fully autonomous. Nonetheless, while countries like the US have vowed to keep the most morally significant final decision within the orbit of human controllers, other states—especially those with authoritarian proclivities—as well as nonstate actors, could develop LAWs that are fully autonomous. This is worrisome in part because current technology does not enable such weapons to reliably discriminate between combatants and noncombatants (Asaro 2008), a significant ethical/legal problem given that the Laws of War (LoW) outlined in the Geneva and Hague conventions prohibit the killing of civilians and surrendering of combatants, except under strict circumstances where the principle of proportionality justifies the possibility of unwanted deaths to prevent further loss. Similarly, there are important questions concerning the sorts of artificial intelligence (AI) systems that should govern the behavior of LAWs, as

well as issues regarding how best to resolve any potential legal complications arising from the use of LAWs in the theater of war.

The present paper will bracket these topics, focusing instead on the possible virtues of using LAWs in battle. To the best of our knowledge, it is the first study to explore the potential *merits* of ethically-programmed LAWs, and to outline several prescriptions that could inform an international regulatory framework for such weapons. In contrast, prior studies of LAWs have focused on their feasibility (Singer 2009b; Thurnher 2016), implications (Guetlein 2005; Krishnan 2009), legal consequences (Lewis 2015; Marchant et al. 2015; Marauhn 2014; Sauer 2016; Sharkey 2010; Thurnher 2013), moral-system actions (Johnson and Axinn 2013), and programmability (Arkin 2008; Danielson 1999; DeBaets 2014; Goertzel 2016; Lin, Bekey, and Abney 2008; Pereira and Saptawijaya 2007; Shulman, Jonsson, and Tarleton 2009; Soares 2016; Tarleton 2010; Wallach, Allen, and Smit 2008). While this literature offers many useful insights about the nature and possible complications of LAWs, it does not adequately address the possible moral upsides of relying on LAWs to make increasingly autonomous critical decisions during conflict rather than human judgment. Indeed, this paper suggests that we should resist a blanket ban on LAWs, instead pursuing moratoria that selectively prohibit the development of particular types of LAWs while simultaneously *encouraging* the development of other types. To make this argument, we will first review some of the scholarship that favors a blanket ban of this emerging technology. We will then provide a detailed discussion of the possible benefits of LAWs and argue that there may be a moral imperative to actively develop *moral machines*—that is, “ethical LAWs.” Finally, we conclude by specifying a recommendation that, we believe, should constitute a critical element of any future treaty for governing the use of LAWs by states.

2. The Argument for Prohibition

Founded in 2013, the Campaign to Ban Killer Robots (KRC) aims to create an international body of legislation to regulate—if not categorically prohibit—the martial use of artificial systems that exclude humans from the decision-making process (Heyns 2013). While KRC and the ethical/legal scholars who are sympathetic with its mission have proposed a range of arguments for prohibiting LAWs, the two central arguments are: (i) introducing LAWs to the battlefield would reduce the threshold for conflict because of their relatively low production cost, social cost, and ease of use; and (ii) current LAWs do not have the technological capacity to accurately categorize valid and invalid targets, thus rendering them dangerously incapable of satisfying the laws of armed conflict. Taking these in order:

2.1 *The threshold for war*

To begin, one of the most compelling reasons for *opposing* nuclear non-proliferation efforts is that the destructive potential of nuclear weapons increases the threshold of use (W. Wilson 2012; Jürgen 2008). Thus, only in extreme circumstances would rational actors deem their use to be either morally or strategically acceptable. This strongly contrasts with the case of LAWs, whose cost would be small compared to the cost of paying military personnel. Consequently, states could maintain stockpiles of LAWs that are far larger than any standing army. The low cost of LAWs would also make them more expendable than human soldiers (Jenks 2010; Jacoby and Chang 2008; P. Singer 2009a), and they could strike the enemy with greater precision than human soldiers can currently achieve (Ekelhof and Struyk 2014; Thurnher 2012). These four properties—low cost, military effectiveness, expendability, and precision—could drive proliferation while *lowering* the threshold for use and, therefore, undermine geopolitical security. Incidentally, similar claims could be made about anticipated future nanotech weaponry (see Whitman 2011).

The attractiveness of LAWs is apparent in the US’s use of “unmanned aerial vehicles” (UAVs, also known as “drones”) in Iraq and Syria. These semi-autonomous systems offer a cheap, effective, and relatively precise means for conducting surveillance and targeting enemy combatants [despite unsatisfied infrastructural needs to sustain the drone program] (McLean 2014). As a result, the US drone program has

grown and the frequency of drone use against terrorist organizations like the (now-defunct) Islamic State has steadily increased in the past decade (Higgins 2017). Yet the proliferation of LAWs discussed in this paper is different in important respects from the proliferation of current UAV devices. Most obviously, LAWs are moral actors capable of making life and death decisions without human intervention. The absence of a human operator suggests that LAWs will be even cheaper than current UAVs and, as such, more vulnerable to proliferation. But this might not be undesirable given that, for example, ethical LAWs will—almost by definition—not serve to glorify or extend war efforts beyond the initial scope. Furthermore, UAVs still require human intervention and, as we will soon discuss, the emotional volatility of humans could lead to a number of otherwise avoidable problems.

More generally speaking, the growing use of UAVs in conflict situations is consistent with a broader trend toward high-precision weaponry and away from larger, more destructive weapons like those in the world's nuclear arsenals (G. Wilson 2013). There are some reasons for welcoming this shift. For example, the use high-precision weapons like LAWs to achieve a state's military objectives could reduce the probability and proportion of indiscriminate harm, thus violating the LoW and "rules of engagement" (RoE) less than might otherwise have been possible. Even more, the "ease-of-use" of LAWs that are fully autonomous could enhance the "balance of terror" that prevents conflict from breaking out by providing a credible means for retaliation: "If you strike me first, I will unleash a swarm of LAWs that devastate your infrastructure, poison your streams, set fire to your farms, destroy your armies, and assassinate your leaders."

The precision and effectiveness of LAWs could also accelerate the process of nuclear disarmament. First, consider that research on the potential climatic consequences of a nuclear war resulted in the replacement of MAD ("mutually-assured destruction") with SAD ("self-assured destruction"). The reason is that an exchange of nuclear weapons—even a regional one [citation]—could initiate a "nuclear winter" that causes global agricultural failures, widespread starvation, the spread of infectious disease, and other catastrophic sequelae that cannot be contained within national borders (Mills et al. 2014; Xia et al. 2015). Consequently, a nuclear war would all but guarantee the *self*-annihilation of states involved. As Seth Baum (2015) notes, though, LAWs could provide a kind of "winter-safe deterrence" by providing states with a credible threat of retaliation without the global catastrophic risks of nuclear conflict. Thus, LAWs could render the world's nuclear arsenals irrelevant and, in doing so, lower the overall risk of human annihilation.

2.2 Distinguishing valid targets

As mentioned above and discussed further below, the most significant obstacle to the development and acceptance of LAWs is their current inability to reliably distinguish between valid and invalid targets. This is a crucial shortcoming because, as previously alluded, the Geneva Conventions require that attacks must be exclusively directed at military rather than civilian entities (United Nations 1979). As Article 51(3) of Section 1 of the Additional Protocol to the Geneva Convention states, "civilians shall enjoy the protection afforded by this Section, unless and for such time as they take a direct part in hostilities" (United Nations 1979).ⁱⁱ Complicating the situation further is an ambiguity about what exactly counts as a "civilian"—e.g., some civilians engage in combative activities, some of which are not clearly *overt*, thus making it a matter of interpretation whether such individuals should fall within the "combatant" or "noncombatant" categories.

Likewise, targeting military activities can sometimes result in harm to non-military agents. Scenarios of this sort are governed by the principle of proportionality, which "requires that the expected harm to civilians be measured, prior to the attack, against the anticipated military advantage to be gained from the operation. This rule described as 'one of the most complex rules of international humanitarian law,' is largely dependent on subjective estimates of value and context-specificity" (Heyns 2013, § 70). The key term here is "estimates of value," since it depends upon some moral decision-making structure for its meaning. Here one can argue that LAWs should not be permitted to make lethal decisions because of their inability not only to accurately categorize the targets that they encounter, but to estimate the value

and understand the context of those targets (de Boisboissel 2015; Roff 2014).ⁱⁱⁱ Thus, KRC and similar organizations argue that humans should remain in the decision-making loop (Heyns 2013)—for example, the International Committee for Robot Arms Control (ICRAC), founded in 2009 by Juergen Altmann, Peter Asaro, Noel Sharkey, and Rob Sparrow, specifies its mission as inspiring the “international community [to create] a legally binding treaty to prohibit the development, testing, production and use of autonomous weapon systems in all circumstances” (ICRAC 2009, 2014)

3. The Argument for LAW Development

3.1 Human flaws and foibles

Yet, we would argue, such positions are predicated on an unfounded fear that taking control away from humans will enable robotic weaponry to demolish current, human-involved warfare practices. Extrapolating techno-development trends into the future, it is reasonable to expect future robotic weapons to acquire the capacity to reliably and accurately differentiate between combatants and noncombatants (Egeland 2016; Sharkey 2012); this could even occur in the near future (see Guizzo 2016). Indeed, Ronald Arkin (2008) anticipates such technologies—in particular, recognition software—to not only be developed but surpass human performance capabilities (see also Egeland 2016; O’Meara 2011). As he writes, “we must protect the innocent non-combatants in the battlespace far better than we currently do. Technology can, must, and should be used toward that end.” Like Nadeau, Arkin believes that moral LAWs would act in an ethically superior way to humans in war, saying that:

The commonplace occurrence of slaughtering civilians in conflict over millennia gives rise to my pessimism in reforming human behaviour yet provides optimism for robots being able to exceed human moral performance in similar circumstances (Arkin 2015)

One must also take into account the consequences of humans personally engaging in warfare. Historical records, including those of concurrent military engagements, recount numerous acts of barbarism as a result of the harsh conditions that combatants are exposed to (Arkin 2015). In fact, Lin et al. (2008) discuss how one of the most attractive prospects of LAWs is their inability to be affected by emotions on the battlefield (Lin, Bekey, and Abney 2008). It is the emotional distress that often causes combatants to mistreat the enemy and commit war crimes. Hence, the introduction of LAWs that are unaffected by such emotional stress serves as an incentive for continued development (Klincewicz 2015).^{iv}

Secondly, the emotional and physical pressures that human combatants must endure during wartime has performance costs. The fatigue of a long and drawn-out battle affects the ability of individual soldiers to perform optimally, and thus affects the accuracy of their shots (Nibbeling et al. 2014; Burke et al. 2007). LAWs are naturally unaffected by similar physical pitfalls and can always--as long as the physical infrastructure is designed optimally from the start--permit the LAWs to continually perform accurately and as expected.

The ability for LAWs to engage in unwavering, precise combat also resolves some ethical issues that arise from human-waged war. In light of the fact that LAWs don’t possess emotions to guide their behaviours or personal stakes that affect their combat approaches, LAWs will always perform duties accurately under even the most physically--or to a human, emotionally--stressful conditions, thus enabling them to, at least more often than not, kill in a more humane manner. LAWs can be programmed to only engage targets in manners deemed most ethical based on the dynamics of war at the time of combat: the changing environment, the weapons being used by both the aggressor and the defender, and the characteristics of the target (human, robot, or physical structure).

Already, computerized weapons platforms can engage targets far more accurately than any human counterpart can (United States Navy 2017; Geibel 1997; Shachtman 2007; Katz and Lappin 2012). Strong arguments can be levied that LAWs outfitted with such weapons platforms could engage in

otherwise normal wartime duties but in a means that is far more accurate and thus ethical as a consequence of LAWs' technological superiority.

Part of this ethical prowess exhibited by LAWs, is not only because they never tire, but because they are impervious to the psychological shortcomings of humans. Though a contentious topic, several high-profile cognitive psychologists suggest that humans fabricate reasons for their actions after committing them (Nadeau 2006; Davidson 1982). Thus, it is human to be irrational, to make unreasoned decisions toward an action that is then validated after carrying through. Such is not the nature of a robot. As mentioned, LAWs don't have any particular affinity to or personal interests in surviving battle; they don't have any drive to exhibit particular harshness against enemies of a certain culture; and they don't, outside of their goals, worry about winning the war and heading back home after using any unsavory methods to do so. What they do mind is their particular set of rules, their value-laden code that dictates how they are to conduct themselves in an ethical manner during combat.

In sum, the two above arguments (1) the lack of an agreed-upon universal moral framework coupled with (2) the emotional and psychological impacts of war on humans and the consequent tragedies and irrational behaviors that follow, provide a strong case *for* the development and utilization of an emotionally uncompromisable artificial moral combatant--a moral LAW.

Up until this point, this paper has presented arguments against and for the development of LAWs. The remainder of this paper will argue that 1) *only* moral LAWs should be used in warfare and 2) all other non-moral LAWs should be prohibited.

3.2 Moral LAWs and the burden of responsibility

The previous section delineated some indisputable reasons why morally programmed LAWs would surpass the ethical capacities of humans by overcoming our physical shortcomings and susceptibility to morally egregious behavior. We will now argue that *only* moral LAWs should be employed as combatants in war. The exclusivity of that statement does not only include non-morally programmed LAWs but extends to humans as well, when speaking of matters related to organized war. To do this, we draw from the work of Joseph E Nadeau in his controversial paper *Only Androids Can be Ethical* (2008).

Here we adopt Nadeau's (2006) thesis as our own: "humans cannot reason logically except in very limited situations." Nadeau proceeds with a discussion of empirical evidence provided by psychological research that concludes, "...however great the human capacity for vision, speech and classification, the bounds of human reason are very low indeed" (Nadeau 2006, 242; see also Chase 2015; Barrat 2013). Similar to our psychological vulnerabilities discussed in the previous section, these cognitive limitations bind us in a way that leads us to act without reasons. Nadeau hence argues an absolute limit on our free will—and thus liability—on account of limited rationality.

Nadeau, accepting these human limitations, moves toward discussing the potential for androids to surpass the rationality of humans, as well as the possibility that LAWs "could have free will in the philosopher's sense, and so be responsible and culpable" (Nadeau 2006). Thus, arguments made toward holding humans liable for specific actions performed during war, especially during particularly threatening situations, become highly controversial, if not totally disregarded because of their convolution. Nadeau argues that an android architecture built on either a theorem prover, neural network, or a combination of the two would be sufficient to allow the system to initiate each and every action in response to reasons embedded in the programming—whether or not we can actually determine the specific reason, as is the case with neural nets (Nadeau 2006; Arkin 2008; Lin, Bekey, and Abney 2008). The argument can thus be formalized as follows:

1. For an agent to be held responsible and liable, that agent must be acting with a free will.
2. Free actions are only free actions if those actions have one or more antecedent reasons behind them.

3. Human actions are not based on antecedent reasons except in very limited and rare circumstances.
4. A robotic architecture that is built using either a theorem prover or neural network (or combination) makes decisions based on antecedent reasons.
5. Therefore, it is *only* androids, not humans, that can act ethically.

The conclusion imposes a moral obligation on us, as lawmakers, ethicists, and society in general. Given our knowledge of the war crimes caused by humans throughout history, innate human irrationality in most cases of action, and that LAWs are currently in development, it behooves us, morally speaking, to not only program moral values and ethical reasoning into LAWs, but employ LAWs as the sole actor in combat settings. This, of course, does not discount current technological limitations such as targeting systems that are incapable of distinguishing combatants from noncombatants. Given that this is the case, we must take into account that even a moral agent equipped with such rudimentary targeting systems would not gather the sufficient data to make a valid decision. The above argument is crucially conditional: it states that *if* we can create technologies capable of distinguishing combatants from noncombatants, then we should program moral values into them; and *if* we can program moral values into them, then *only* these technologies should be considered ethical, and thus used as combatants in war.

Nonetheless, in lieu of such targeting technology, we should look for a governance framework that is *precautionary* as well as *anticipatory*. International bodies must understand the current limits of technology and see that LAWs are being developed regardless of their interest in researching such developments themselves or employing the technology in their future war efforts and policies. Hence, precautions must be taken when creating laws to govern the development and use of LAWs, on account of the uncertainties that exist regarding how such laws will affect society. Although comprehensive, the framework must also be flexible enough to allow for harmonization in domestic regulatory statutes as well as permit change as technology evolves. The following section discusses the regulatory sphere governing LAWs.

4. Moral and Regulatory Frameworks

4.1 Moral disagreement and regulatory frameworks

In the first part of this paper, we discussed some of the main reasons proposed for why the development of LAWs should either be prohibited entirely or regulated to offer some guidance to progress. One of the most impactful arguments for prohibiting LAW's is their inability to make valid target choices and its ability to take the control of life and death away from human operators by means of exerting its own decision-making processes. However, when one looks at the question of what constitutes a valid target, they are in essence making a judgment call of whether it is permissible to attack. Permissibility, regardless of whether or not it is being considered under frameworks such as the LoW or the RoE, is fundamentally an ethical/moral decision. Thus, two questions surface (1) is any moral framework necessary to be programmed into a law that goes beyond the LoW and RoE? and (2) if so, under which moral framework does the decision maker decide that a target's right to life is suspended and warrants elimination?

Currently, there is no universally agreed upon moral framework. Divine Command Theory (DCT) seeks to dictate correct (i.e., moral) action in accordance with the word of the divine (i.e., holy scripture). Utilitarianism weighs the welfare of varying options presented and determine the best consequences for the greatest number. Deontology, in contrast to the consequentialist aspects of utilitarianism, looks to apply universal moral principles to dictate action. And finally, virtue ethics asks agents to embody characteristics of ideal moral characters (e.g., Christ, Socrates, Aurelius, and Epictetus) and base our actions accordingly. Each of these, however, encounters nontrivial problems. DCT is highly interpretive and, consequently, differences in interpretations of *which* words of *which* divine being quickly arise, with no good epistemic basis for arbitrating between alternatives. Utilitarianism engenders problems associated

with what constitutes the good, which consequences of an action are morally relevant, and how to weigh and calculate the amount of good brought about by an action. Likewise, deontology sometimes yields unpalatable consequences, such as being duty-bound to tell truths with obviously negative repercussions (e.g., telling a Nazi soldier that one is hiding an escapee in one's basement) (Varden 2010).

Given ongoing debate about which moral systems are best—indeed, about 25 percent of philosophers are deontologists, 23 percent are consequentialists, and 18 percent are virtue ethicists (Bourget and Chalmers 2013)—as well as the problems inherent in different moral frameworks, it is difficult to argue that any human has the moral authority to make definitive decisions relating to the right to life of other, even in combat zones. This raises questions of whether or not LAWs, or AI in general, are any different to a human in their ability to interpret and apply moral principles. Yet some scholars argue, as alluded to above, that it could be *only* artificial minds that *can* truly be ethical. This issues make the decision incredibly difficult and would most likely result in a lack of global agreement. Regardless, although some may argue that removing the human from the battlefield and replacing him/her with a robotic counterpart will strip morality in general from war, it is not obvious that such is even necessary. Strong arguments can be made in favour of programming exclusively the LoW and RoE and nothing beyond. Not only would such make international regulation of LAWs simpler, given that the LoW are already established, but it would remove the burden of value-laden programming from the developers of different LAWs manufacturers as well as the difficulty of establishing a universal theory of morality that would be accepted by all lawmakers.

Thus, although morality is per se a contentious topic, particularly when attempting to find a universal moral foundation, such should not come into discussions regarding programming LAWs. Not only does it create new programming hurdles, thus incentivizing the production of LAWs without value programming, but morals beyond the LoW appear unnecessary within the context of war itself.

Likewise, we argue that an international governance framework should take this proposition seriously. The current LoW govern human conduct in war and punish war crimes. Naturally, they exist because such crimes are anticipated and historically have always been committed in such events. However, the applicability of the LoW and national RoE can become solidified as a *intrinsic* part of the behavioral repertoire of moral LAWs. Hence, we argue that an international regulatory framework should be assembled to govern the use of autonomous robots in war and that this framework should include the following considerations:

1. Insofar as weapons are autonomous, ie., beyond human control, they must *necessarily* have an agreed upon moral programming.
2. If states produce moral LAWs of this sort, those LAWs must *always* have the capacity to make judgment calls equal to or greater than humans, e.g., using sensors that can gather combatant data equivalent to or beyond that of humans.

The exclusivity of the above two conditions thus prohibits the use of any LAW that does not have (i) moral programming, and (ii) adequate targeting systems. It follows that international regulations should be created to incorporate those prohibitions. The conditions, though, are not jointly exhaustive, as there are many additional issues that legislators must address. Rather, they are intended to form part of a mandatory standard that constitutes only one piece of a larger, more inclusive framework of regulatory prescriptions (or proscriptions). Further scholarly work is needed to clarify the nuances this issue and convince skeptics to endorse LAWs as an ethical alternative to the contemporary warfare status quo.

4.2 Convention of certain conventional weapons (CCW) and regulation

Furthermore, the 1980 United Nations CCW consists of five protocols which prohibit the use of certain forms of conventional munitions. These include (I) non-detectable fragments, (II) mines and similar trap devices, (III) incendiary weapons, (IV) blinding laser weapons and (V) explosive remnants of

war (ICRC 2014). A similar protocol could be introduced that prohibits LAWs that fall within the exclusivity boundaries of the first two conditions. As mentioned above, a blanket ban on LAWs—although often argued with good intentions—misses certain key moral advantages that ethical LAWs with sufficiently accurate targeting systems would embody. Hence, a better solution would be to endorse *selective regulation and prohibition* to ensure that the research, development, and use of LAWs yields not just morally acceptable outcomes, but morally better ones.

Additionally, states should research and development LAWs in a fully open and transparent manner. Although conducting secretive research and development has both tactical and technological advantages for states in the anarchic realm of international politics, embracing openness and transparency is more conducive to *scientific progress*. As Kantrowitz (1992) puts it,

openness is necessary for the processes of trial and the elimination of error. ... Impediments to the elimination of errors will determine the pace of progress in science as they do in many other matters. ... Trial and the elimination of error is essential to significant progress in military technology, and thus both aspects of the process by which significant progress is made in military technology are sharply decelerated when secrecy is widespread in peacetime. Openness accelerates progress. In peacetime military technology, openness is a weapon. It is one clue to the survival of open societies in an international jungle (Kantrowitz 1992).

The endeavor toward transparency, which Kantrowitz also discusses, can and ought to be applied to LAWs. This could help ensure that (a) states develop them in accordance with internationally agreed upon regulatory statutes, and (b) ethical progress occurs with respect to the science and engineering aspects of this emerging technology. Indeed, in his 2013 report on LAWs, the UN Special Rapporteur Christof Heyns calls for “full transparency regarding all aspects of the development of robotic weapon systems,” thus acknowledging the benefits, if not the necessity, of the transparent development of military technologies (Heyns 2013).

Similarly, the novelty of LAWS as a weapons platform can clearly fall under the domain of Article 36 of the Geneva Conventions Protocol I which explores the process of the “study, development, acquisition or adoption of a new weapon, means or method of warfare.” Although Lt. Col. Justin McClelland has explored this protocol in depth and explicated its governance scope, the long term inclusion of LAWS as fully implicated under Article 36 is definitely a worthwhile place to start as concerns LAWS and a fruitful potential research avenue for scholars to explore (McClelland 2003).

5. Conclusion

This paper has surveyed various arguments both for and against the prohibition of LAWs. On the one hand, some scholars argue that inadequate targeting systems and the removal of humans from the kill control are sufficient for banning LAWs. On the other hand, we contend that the relatively low cost of LAWs, their potential for moral programming, and their ability to remove human combatants from the line of fire constitute strong reasons for pursuing the development and use of LAWs in conflict situations. Even more, we argue that “moral LAWs” could constitute the *only* entities capable of making genuinely ethical decisions about whether its targets live or die. But here we encounter a number of caveats. First, such LAWs must have targeting and judgment systems that are equal or superior to the targeting abilities of humans. Second, it must embody a moral program or programs that all parties agree upon, such as the LoW and RoE. And third, states that ignore these requirements will be subject to international pressure and sanctions. We also suggest that a new protocol that references LAWs could be added to the CCW.

Although we have tried our best to distill these complicated issues and address the relevant problems with our proposal, many issues regarding the creation and employment of LAWs remain unclear. Nonetheless, certain emerging work in the fields of machine ethics and machine learning will increasingly reveal how the future could unfold. The best way forward at this time is to foment

international discussion among researchers and policymakers about these topics; we hope the present paper has made a small contribution toward this end.

References

- Arkin, Ronald C. 2008. "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/reactive Robot Architecture Part I: Motivation and Philosophy." In *Proceedings of the 3rd International Conference on Human Robot Interaction - HRI '08*, 121. New York, New York, USA: ACM Press. doi:10.1145/1349822.1349839.
- . 2015. "The Case for Banning Killer Robots." *Communications of the ACM* 58 (12): 46–47. doi:10.1145/2835965.
- Asaro, Peter M. 2008. "How Just Could a Robot War Be?" In *Proceedings of the 2008 Conference on Current Issues in Computing and Philosophy*, 50–64. Amsterdam, The Netherlands, The Netherlands: IOS Press. <http://dl.acm.org/citation.cfm?id=1566234.1566243>.
- Barrat, James. 2013. *Our Final Invention*. New York, NY: Thomas Dunne Books.
- Baum, Seth D. 2015. "Winter-Safe Deterrence: The Risk of Nuclear Winter and Its Challenge to Deterrence." *Contemporary Security Policy* 36 (1). Taylor & Francis: 123–48. doi:10.1080/13523260.2015.1012346.
- Boisboissel, Gerard de. 2015. "Uses of Lethal Autonomous Weapon Systems." In *International Conference on Military Technologies (ICMT) 2015*, 1–6. IEEE. doi:10.1109/MILTECHS.2015.7153656.
- Bourget, David, and David Chalmers. 2013. What Do Philosophers Believe? *Philosophical Studies*. 170(3): 465-500.
- Burke, Kelly A, Tal Oron-Gilad, Gareth Conway, and Peter A Hancock. 2007. "Friend/Foe Identification and Shooting Performance: Effects of Prior Task Loading and Time Pressure." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 51 (4). SAGE Publications Inc: 156–60. doi:10.1177/154193120705100403.
- Chase, Calum. 2015. *Surviving AI: The Promise and Peril of Artificial Intelligence*. London: Three Cs.
- Danielson, Peter. 1999. "Evolutionary Models of Cooperative Mechanisms: Artificial Morality and Genetic Programming." In *Modeling Rationality, Morality, and Evolution*, edited by Peter Danielson, 423–62. Oxford University Press. <https://global.oup.com/academic/product/modeling-rationality-morality-and-evolution-9780195125498?cc=ca&lang=en&>.
- Davidson, Donald. 1982. "Paradoxes of Irrationality." In *Problems of Rationality*, 189–98. Oxford University Press. <https://academiaanalitica.files.wordpress.com/2016/10/donald-davidson-problems-of-rationality-2004.pdf>.
- DeBaets, Amy Michelle. 2014. "Can a Robot Pursue the Good? Exploring Artificial Moral Agency." *Journal of Evolution and Technology* 24 (3): 76–86. <http://jetpress.org/v24.3/DeBaets.htm>.
- Egeland, Kjøl. 2016. "Lethal Autonomous Weapon Systems under International Humanitarian Law." *Nordic Journal of International Law* 85 (2): 89–118. doi:10.1163/15718107-08502001.
- Ekelhof, Merel, and Miriam Struyk. 2014. "Deadly Decisions: 8 Objections to Killer Robots - Google Books." https://books.google.ca/books/about/Deadly_Decisions.html?id=66UXrgEACAAJ&redir_esc=y.
- Geibel, Adam. 1997. "Learning from Their Mistakes: Russia's Arena Active Protection System." <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA323300>.
- Geiss, Robin. 2015. "The International-Law Dimension of Autonomous Weapons Systems." Friedrich-Ebert-Stiftung, International Policy Analysis. <http://eprints.gla.ac.uk/117554/>.
- Germany. 2014. "General Statement by Germany, CCW Expert Meeting Lethal Autonomous Weapons Systems." Geneva. [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/97636DEC6F1CBF56C1257E26005FE337/\\$file/2015_LAWS_MX_Germany.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/97636DEC6F1CBF56C1257E26005FE337/$file/2015_LAWS_MX_Germany.pdf).
- Goertzel, Ben. 2016. "Infusing Advanced AGIs with Human-Like Value Systems : Two Theses." *Journal*

- of Evolution and Technology* 26 (1): 50–72.
- Guetlein, Michael A. 2005. “Lethal Autonomous Weapons -- Ethical and Doctrinal Implications.” Newport: Naval War College. <http://www.dtic.mil/docs/citations/ADA464896>.
- Guizzo, Erico. 2016. “Autonomous Weapons ‘Could Be Developed for Use Within Years,’ Says Arms-Control Group.” *IEEE Spectrum*. <https://spectrum.ieee.org/autaton/robotics/military-robots/autonomous-weapons-could-be-developed-for-use-within-years>.
- Heyns, Christof. 2013. *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. Human Rights Council*. United Nations General Assembly. http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.
- Higgins, Eoin. 2017. “Obama’s Drone War Escalating in His Last Year in Office.” *Huffington Post*. http://www.huffingtonpost.com/entry/obamas-drone-war-escalating_b_9611490.html.
- ICRAC. 2009. “Original Mission Statement.” *International Committee for Robot Arms Control*. <https://icrac.net/statements/>.
- . 2014. “2014 Mission Statement.” *International Committee for Robot Arms Control*. <https://icrac.net/statements/>.
- ICRC. 2014. “1980 Convention on Certain Conventional Weapons – Factsheet.” *International Committee of the Red Cross*. <https://www.icrc.org/en/document/1980-convention-certain-conventional-weapons>.
- Jacoby, Grant A., and David J. Chang. 2008. “Towards Command and Control Networking of Cooperative Autonomous Robotics for Military Applications (CARMA).” In *2008 Canadian Conference on Electrical and Computer Engineering*, 000815–20. IEEE. doi:10.1109/CCECE.2008.4564649.
- Jenks, Chris. 2010. “Law from Above: Unmanned Aerial Systems, Use of Force, and the Law of Armed Conflict.” *North Dakota Law Review*, 85: 649. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1569904.
- Johnson, Aaron M., and Sidney Axinn. 2013. “The Morality of Autonomous Robots.” *Journal of Military Ethics* 12 (2): 129–41. doi:10.1080/15027570.2013.818399.
- Jürgen, Altmann. 2008. “Military Uses of Nanotechnology - Too Much Complexity for International Security?” *Complexity* 2 (1): 62–70. doi:10.1002/cplx.
- Kantrowitz, Arthur. 1992. “The Weapon of Openness.” In *Nanotechnology Research and Perspectives*, edited by B. C. Crandal and James Lewis, 303–11. Cambridge, MA: MIT Press. <https://www.foresight.org/Updates/Background4.html>.
- Katz, Yaakov, and Yaakov Lappin. 2012. “Iron Dome Ups Its Interception Rate to over 90%.” *Jerusalem Post*. <http://www.jpost.com/Defense/Iron-Dome-ups-its-interception-rate-to-over-90-percent>.
- Kirk, Ashley. 2015. “What Are the Biggest Defence Budgets in the World? - Telegraph.” *The Telegraph*. <http://www.telegraph.co.uk/news/uknews/defence/11936179/What-are-the-biggest-defence-budgets-in-the-world.html>.
- Klincewicz, Michał. 2015. “Autonomous Weapons Systems, the Frame Problem and Computer Security.” *Journal of Military Ethics* 14 (2). Taylor & Francis: 162–76. doi:10.1080/15027570.2015.1069013.
- Krishnan, Armin. 2009. *Killer Robots : Legality and Ethicality of Autonomous Weapons*. Ashgate. https://books.google.ca/books/about/Killer_Robots.html?id=klvIN9PgBeYC.
- Lewis, John. 2015. “The Case for Regulating Fully Autonomous Weapons.” *Yale L.J.* 1309, January. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2528370.
- Lin, Patrick George, Keith Bekey, and M A Abney. 2008. “Autonomous Military Robotics: Risk, Ethics, and Design.” http://ethics.calpoly.edu/onr_report.pdf.
- Marauhn, Thilo. 2014. “An Analysis of the Potential Impact of Lethal Autonomous Weapons Systems on Responsibility and Accountability for Violations of International Law.” In *CCW Expert Meeting on Lethal Autonomous Systems*. Geneva. https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/assets/media/35FEA015C2466A57C1257CE4004BCA51/file/Marauhn_MX_Laws_SpeakingNotes_2014.pdf.

- Marchant, Gary E, Braden Allenby, Ronald C Arkin, Jason Borenstein, Lyn M Gaudet, Orde Kittrie, Patrick Lin, George R Lucas, Richard O'Meara, and Jared Silberman. 2015. "International Governance of Autonomous Military Robots." In *Handbook of Unmanned Aerial Vehicles*, edited by Kimon P Valavanis and George J Vachtsevanos, 2879–2910. Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-9707-1_102.
- McClelland, Justin. 2003. "The Review of Weapons in Accordance with Article 36 of Additional Protocol I." https://www.icrc.org/eng/assets/files/other/irrc_850_mcclelland.pdf.
- McLean, Wayne. 2014. "Drones Are Cheap, Soldiers Are Not: A Cost-Benefit Analysis of War." *The Conversation*. <http://theconversation.com/drones-are-cheap-soldiers-are-not-a-cost-benefit-analysis-of-war-27924>.
- Mills, Michael J., Owen B. Toon, Julia Lee-Taylor, and Alan Robock. 2014. "Multidecadal Global Cooling and Unprecedented Ozone Loss Following a Regional Nuclear Conflict." *Earth's Future* 2 (4). Wiley Periodicals, Inc.: 161–76. doi:10.1002/2013EF000205.
- Nadeau, Joseph E. 2006. "Only Androids Can Be Ethical." In *Thinking about Android Epistemology*, edited by K. M. Ford, C. N. Glymour, and P. J. Hayes. AAAI Press (American Association for Artificial Intelligence).
- Nibbeling, Nicky, Raoul R.D. Oudejans, Emiel M. Ubink, and Hein A.M. Daanen. 2014. "The Effects of Anxiety and Exercise-Induced Fatigue on Shooting Accuracy and Cognitive Performance in Infantry Soldiers." *Ergonomics* 57 (9): 1366–79. doi:10.1080/00140139.2014.924572.
- O'Meara, Richard M. 2011. "Contemporary Governance Architecture Regarding Robotics Technologies: An Assessment." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Bekey. George A., Keith Abney, and Patrick Lin, 159–68. Cambridge, MA: MIT Press. <http://arteca.mit.edu/book/robot-ethics>.
- Pereira, Luís Moniz, and Ari Saptawijaya. 2007. "Modelling Morality with Prospective Logic BT - Progress in Artificial Intelligence: 13th Portuguese Conference on Artificial Intelligence, EPIA 2007, Workshops: GAIW, AIASTS, ALEA, AMITA, BAOSW, BI, CMBSB, IROBOT, MASTA, STCS, and TEMA, Guimarães, Portug." In , edited by José Neves, Manuel Filipe Santos, and José Manuel Machado, 99–111. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-77002-2_9.
- Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York, NY: Viking Books.
- Roff, Heather M. 2014. "The Strategic Robot Problem: Lethal Autonomous Weapons in War." *Journal of Military Ethics* 13 (3). Taylor & Francis: 211–27. doi:10.1080/15027570.2014.975010.
- Sauer, Frank. 2016. "Stopping 'Killer Robots': Why Now Is the Time to Ban Autonomous Weapons Systems." *Arms Control Association*. https://www.armscontrol.org/ACT/2016_10/Features/Stopping-Killer-Robots-Why-Now-Is-the-Time-to-Ban-Autonomous-Weapons-Systems.
- Shachtman, Noah. 2007. "Robo-Snipers, 'Auto Kill Zones' to Protect Israeli Border." *Wired*. https://www.wired.com/2007/06/for_years_and_y/.
- Sharkey, Noel E. 2010. "Saying 'No!' to Lethal Autonomous Targeting." *Journal of Military Ethics* 9 (4). Routledge: 369–83. doi:10.1080/15027570.2010.537903.
- . 2012. "The Evitability of Autonomous Robot Warfare." *The Evitability of Autonomous Robot Warfare* 886: 787–99. <https://www.icrc.org/eng/resources/documents/article/review-2012/irrc-886-sharkey.htm>.
- Shulman, Carl, Henrik Jonsson, and Nick Tarleton. 2009. "Machine Ethics and Superintelligence." In *AP-CAP 2009: The Fifth Asia-Pacific Computing and Philosophy Conference*, 95–97. Tokyo. <http://ia-cap.org/ap-cap09/proceedings.pdf>.
- Singer, Peter W. 2009a. "Military Robots and the Laws of War." *The New Atlantis*. <http://cvrr.ucsd.edu/ece172a/fa10/papers/MilitaryRobotsSinger2009.pdf>.
- . 2009b. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. Edited by null. Vol. null. Null.

- Soares, Nate. 2016. "The Value Learning Problem." In *Ethics for Artificial Intelligence Workshop at 25th International Joint Conference on Artificial Intelligence*, 1–8. Machine Intelligence Research Institute.
- Tarleton, Nick. 2010. "Coherent Extrapolated Volition: A Meta-Level Approach to Machine Ethics." <https://intelligence.org/files/CEV-MachineEthics.pdf>.
- Thurnher, Jeffrey S. 2012. "Legal Implications of Fully Autonomous Targeting." *Joint Force Quarterly*, no. 67: 77–84. http://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-67/JFQ-67_77-84_Thurnher.pdf.
- . 2013. "The Law That Applies to Autonomous Weapon Systems." *17 ASIL INSIGHTS*, no. 4 (January). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2296343.
- . 2016. "Means and Methods of the Future: Autonomous Systems BT - Targeting: The Challenges of Modern Warfare." In *Targeting: The Challenges of Modern Warfare*, edited by Paul A L Ducheine, Michael N Schmitt, and Frans P B Osinga, 177–99. The Hague: T.M.C. Asser Press. doi:10.1007/978-94-6265-072-5_9.
- UK Ministry of Defence. 2017. "Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems." https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/640299/20170706_JDP_0-30.2_final_CM_web.pdf.
- United Nations. 1979. *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (AP I)*. Vol. 1125. https://treaties.un.org/doc/publication/unts/volume_1125/volume-1125-i-17512-english.pdf.
- United States Navy. 2017. "MK 15 - Phalanx Close-In Weapons System (CIWS)." *The US Navy -- Fact File*. http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2.
- Varden, Helga. 2010. "Kant and Lying to the Murderer at the Door...One More Time: Kant's Legal Philosophy and Lies to Murderers and Nazis." *Journal of Social Philosophy* 41 (4). Blackwell Publishing Inc: 403–21. doi:10.1111/j.1467-9833.2010.01507.x.
- Wallach, Wendell, Colin Allen, and Iva Smit. 2008. "Machine Morality: Bottom-up and Top-down Approaches for Modelling Human Moral Faculties." *AI and Society* 22 (4): 565–82. doi:10.1007/s00146-007-0099-0.
- Walzer, Michael. 1991. *Just and Unjust Wars : A Moral Argument with Historical Illustrations*. https://books.google.ca/books/about/Just_and_Unjust_Wars.html?id=EuTQCQAAQBAJ&redir_esc=y.
- Whitman, Jim. 2011. "The Arms Control Challenges of Nanotechnology." *Contemporary Security Policy* 32 (1). Routledge: 99–115. doi:10.1080/13523260.2011.556848.
- Wilson, Grant. 2013. "Minimizing Global Catastrophic and Existential Risks from Emerging Technologies Through International Law." *Virginia Environmental Law Journal* 31 (1): 307–64. <http://web.a.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10455183&AN=101117008&h=HCtN0Qwj9FkMi0BNk3BVUMyUEAZw1qOJJEHFAYjkQwVTp%2FmyBP2O2XrD%2F3Vl6hfhLdIt91ov6ATK5PhtMHBKA%3D%3D&crl=c&resultNs=AdminWebAuth&resultLocal>.
- Wilson, Ward. 2012. "The Pennyfarthing H-Bomb." *The World Today*. https://www.academia.edu/5105513/Pennyfarthing_H-Bomb.
- Work, Bob. 2015. "The Third U.S. Offset Strategy and Its Implications for Partners and Allies > U.S. DEPARTMENT OF DEFENSE." *U.S. Department of Defense*. <https://www.defense.gov/News/Speeches/Speech-View/Article/606641/the-third-us-offset-strategy-and-its-implications-for-partners-and-allies/>.
- Xia, Lili, Alan Robock, Michael Mills, Andrea Stenke, and Ira Helfand. 2015. "Decadal Reduction of Chinese Agriculture after a Regional Nuclear War." *Earth's Future* 3 (2). Wiley Periodicals, Inc.: 37–48. doi:10.1002/2014EF000283.

ⁱ See Pinker 2011.

ⁱⁱ Herein referred to as AP I.

ⁱⁱⁱ Although some issues within the command and control infrastructure can arise from such an abdication of strategic targeting to LAWS, the philosophical issues at play in the paper remain unaffected given the clearly demarcated philosophical approach taken. Technical and legislative issue to address this must obviously take precedence when aiming to resolve these issue. For a more in depth discussion of these issue, see Roff 2014.

^{iv} Michal Klincewicz (2015) provides a uniquely thorough account of the psychological differentiation between autonomous weapons systems and humans.