

“Old Laws and New LAWS: Just War Theory’s Response to Killer Robots”

Date: 2 September 2019

Word count: 2947

An open letter requesting ‘a ban on offensive autonomous weapons beyond meaningful human control’, has attracted the endorsement of Stephen Hawking, Elon Musk, Apple co-founder Steve Wozniak and Twitter CEO Jack Dorsey, and the signatures of over four thousand academics, since its announcement in July 2015.¹ The letter highlights the risks attached to Lethal Autonomous Weapon Systems (“LAWS”), defined in one instance as weapon systems ‘that, based on conclusions derived from gathered information and preprogramed constraints, is capable of independently selecting and engaging targets’.² Fully autonomous weapons do not currently exist; however this is expected to change as they supplant manned and remote-controlled systems.³ In light of this concern, this essay will examine the extent of regulatory disconnection between the ethical and legal challenges posed by LAWS use and the existing framework of the rules of war. I will first explore the latter’s interaction with historical innovations in military technology, and then analyse whether such approaches are suited to the advent of autonomous weaponry.

1 LEGAL BACKGROUND

There is no one source for the rules of war – that is, those which apply during armed conflict to protect non-participants and regulate the conduct of hostilities.⁴ Like the classic brocards of criminal and civil law, they have emerged from multiple sources both codified and customary, variably referred to as the Laws of Armed Conflict (LOAC) or International Humanitarian Law (IHL).⁵ The most authoritative sources are the Geneva Conventions and its Additional Protocols (ratified in New Zealand through the Geneva Conventions Act 1958). These are supplemented by “just war theory”,⁶ divided by Walzer into *jus ad bellum* (concerning the reasons states have for

¹ Autonomous Weapons: An Open Letter From AI & Robotics Researchers” *Future of Life Institute* <<https://futureoflife.org/open-letter-autonomous-weapons/>>.

² Rebecca Crootof “The Killer Robots Are Here: Legal and Policy Implications” (2015) 36 *Cardozo L. Rev.* 1837 at 1837.

³ Jeffrey S. Thurner “Examining Autonomous Weapon Systems from a Law of Armed Conflict Perspective” in McLaughlin and Nasu *New Technologies and the Law of Armed Conflict* (T.M.C. Asser Press, The Hague, 2014) at 214.

⁴ International Committee of the Red Cross Geneva [“ICRC”] “A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977” (2006) 88 *International Review of the Red Cross* 864 at 932.

⁵ P. W. Singer *Wired For War: The Robotics Revolution and Conflict in the Twenty-first Century* (The Penguin Books, New York, 2009) at 383

⁶ Jai Galliot *Military Robots: Mapping the Moral Landscape* (Ashgate Publishing Ltd, Surrey, 2015) at 71.

going to war) and *jus in bello* (the means states employ within wars).⁷ This essay will concern *jus in bello*'s principles of discrimination and proportionality.⁸ Discrimination requires combatants to distinguish appropriate and inappropriate targets of military force;⁹ namely, differentiating threatening individuals (i.e. 'armed individuals that are aggressively pursuing you') from *hors de combats*, non-combatants and civilians.¹⁰ Galliot notes how the principle of discrimination imposes two challenges to combatants: firstly, the difficulty of distinguishing threatening and non-threatening parties, especially potent in urban environments and with the prevalence of plainclothes insurgent movements.¹¹ The second difficulty is the risk of collateral damage to non-combatants living in close proximity to combat areas. Meanwhile, the proportionality principle requires '*the anticipated moral cost of fighting a war not be out of line with the expected moral benefits*';¹² in Singer's words, the suffering and devastation caused by the war can't outweigh the harm which caused it. Hence, it is not proportionate to respond to a slap in the face by bombing cities.¹³ Galliot takes this principle as guided by consequentialist ethics: the aim of combatants should be to minimise necessary evils and maximise good outcomes. Discrimination and proportionality serve as interpretative aids in the necessary policy debates about regulating emerging technologies. Whether LOAC may effectively encompass the challenges posed by LAWS may therefore depend on whether just war principles and related IHL satisfies regulatory fit and connection.

2 HISTORICAL RESPONSE

We now consider how IHL has evolved in response to past development of weapons of war. Just war theory has adapted to and guided novel forms of conflict; Galliot notes that in the face of modern weapons '*inconceivable for classical just war theorists*', such as chemical, biological and nuclear weaponry, these principles continued to offer practical guidance on the

⁷ Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 2006) at 21.

⁸ Singer, above n 5, at 383.

⁹ Duane Cady *From Warism to Pacifism: A Moral Continuum* (Temple University Press, Philadelphia, 2010) at 37.

¹⁰ Galliot, above n 6, at 78.

¹¹ *Ibid.*

¹² At 79.

¹³ Singer, above n 5, at 383.

moral problems of such warfare”.¹⁴ In spite of the moral deliberation required for its effective use, it is open to conclude that just war theory has remained relevant in the face of emerging military technologies without great need for reform.

That aside, international responses have increasingly followed a downstream regulatory approach. Chemical weapons were not banned until eighty-two years after their initial deployment,¹⁵ considerably greater time than it took a century earlier to ban expanding bullets following their introduction;¹⁶ air warfare was also not addressed until 1977.¹⁷ Additionally, interpretation and adaptation of existing rules may also stagnate: for instance, the 1899 Hague Declaration II on asphyxiating gases did not prevent later gas warfare in World War One.¹⁸ Singer suggests that the biggest catalysts for meaningful legal reform are catastrophes; for instance, the precipitation of the Geneva Conventions by the atrocities of World War Two. This increasingly belated response to technological evolution may be explained as the result of structural/institutional changes: the international table has admitted more players than existed in the 19th century, which may make reaching consensus on universal standards more difficult. Alternatively, there may be political factors at play as well, namely the divided interests of state actors. It would be difficult in the context of the arms races of the 19th and 20th centuries to advocate for restraint in research and development of new weapons. Hence Singer’s exemplar of the introduction of technologies such as bomber planes, chemical weapons and Q-boats in World War One before states could agree on their rules of use.¹⁹ Of note is the present resistance to general regulation of new military technologies by operators, advisers and lawyers, in favour of the enduring general principles’ capacity to encapsulate novel technologies.²⁰ A brief historical analysis seems to favour the notion that regulation of emerging military technologies has experienced a trend of ‘stickiness’ wherein reforms often react to negative consequences of disconnection between the law and reality.

¹⁴ Galliot, above n 6, at 82-83.

¹⁵ Singer, above n 5, at 387.

¹⁶ Robert McLaughlin and Hitoshi Nasu “Introduction: Conundrum of New Technologies in the Law of Armed Conflict” in McLaughlin and Nasu, above n 3, at 4.

¹⁷ William H. Boothby “The Legal Challenges of New Technologies: An Overview” in McLaughlin and Nasu, above n 15, at 26.

¹⁸ Ibid.

¹⁹ Singer, above n 5, at 387.

²⁰ McLaughlin and Nasu, above n 15, at 5-6.

Perhaps in response to this trend, the Geneva Conventions provided through Additional Protocol I, Article 36, a means to their own prospectivity, by requiring contracting states to carry out legal reviews of new weapons and ensure ongoing compliance with both the Conventions and customary elements of weapons law.²¹ According to the International Committee of the Red Cross (ICRC), the sovereign legislator of IHL, Article 36 prevents illegal use of weapons through determining their lawfulness before their development or acquisition.²² This provision is notable because it enabled military regulation to be administered by legal advisers *within* sovereign governments: ongoing legal development can therefore be achieved without the need for the difficult reform of international legislation. In the context of LAWS, many states have already taken measures to ensure compliance with LOAC.²³ However, Article 36 has two problems. The first is political: by entrusting states to hold themselves to account in administering Article 36, that administration may become vulnerable to states' interests in developing certain technologies. The second problem is interpretative: the legality of a particular technology is assessed according to the principles of IHL – namely, just war theory.²⁴ This is not a problem with clear cases of legality or illegality; it may often be easy to determine whether a particular use of a new technology would violate *jus in bello*. But the usefulness of Article 36 as a tool of direct military regulation arguably ends in marginal cases, where just war theory is unable to give a clear answer.

3 PRESENT CHALLENGES

Because quick, prospective or prohibitive international regulation seems rare, and because day-to-day legal responses to emerging technologies are largely left to the principles of just war theory, this inquiry will now consider whether those principles can effectively respond to LAWS. LAWS offer many advantages in warfare, including speed, agility, accuracy, persistence, reach,

²¹ Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts 1125 UNTS 3 (signed 27 November 1978, entered into force 7 December 1979), art. 36.

²² ICRC, above n 4, at 933.

²³ Thurner, above n 3, at 215.

²⁴ *Ibid.*

coordination and mass;²⁵ a popular humanitarian benefit is the removal of the operating power's soldiers from the immediate battlespace and the consequent preservation of their safety.²⁶ The extent to which this is actually the case has been contested,²⁷ and the advantage is nonetheless presently enjoyed by operators of remotely-piloted UAVs, which do not engage targets autonomously; however, Coleman argues that greater future autonomy is likely given the reduced cost of a single operator of multiple machines coupled with the impossibility of that operator keeping up with the simultaneous targeting decisions of each.²⁸ Thurner also mentions the added protection from communications jamming and cyber-attacks when LAWS are not dependent on "tethered links" to operators.²⁹ The benefit of removing soldiers from the battlespace should therefore be weighed against the costs of that removal. Three challenges of LAWS are the dehumanising effect of distance, the problem of marginal cases, and the effect on accountability for war crimes.

The first problem is practical: the fear that '*turning killing into merely the elimination of icons on a computer screen might make the experience feel the same way*' as, in the words of convicted war criminal Steven Green, '*squashing an ant*'.³⁰ Singer argues that physical and emotional distance from the enemy offered by LAWS removes the deterring effects of the up-close horror of battle and the "bond of mutuality" that soldiers and foes gain from their shared few minutes of danger;³¹ while LAWS may remove emotion from the situation, reducing the risk of '*unplanned rage-filled war crimes*' they might also enable '*the type of deliberate war crimes that a professional soldier might refuse*': for instance, an infantryman with a flamethrower or a knife might not have the stomach for the horrors wreaked from afar by the fire-bombers of Hamburg, Dresden and Tokyo.³² Despite this, some argue that the removal of risk to soldiers is a trite exercise in discrimination and proportionality, making just war theory applicable.³³ It is hard to imagine the

²⁵ Vincent Boulanin and Maaïke Verbruggen "Article 36 Reviews: Dealing with the challenges posed by emerging technologies" *Stockholm International Peace Research Institute* (December 2017) <www.sipri.org/sites/default/files/2017-12/article_36_report_1712.pdf> at 19.

²⁶ Galliot, above n 6, at 95-96.

²⁷ At 97-98.

²⁸ Stephen Coleman "Ethical Challenges of New Technologies" in McLaughlin and Nasu, above n 15, at 39 to 40.

²⁹ Thurner, above n 3, at 217.

³⁰ Singer, above n 5, at 395.

³¹ At 396.

³² Galliot, above n 6, at 140.

³³ Daniel Brunstetter and Megan Brun, "The Implications of Drones on the Just War Tradition", (2011) *Ethics and International Affairs* 25(3) at 339.

law requiring that to obviate the risks of emotional distance, soldiers be sent into harm's way; indeed, few thinkers have proposed any legal response to the issue. In my view, however, factors affecting the propensity of actors to use military force are relevant policy considerations in the regulatory process. Promoting, even indirectly, a lower threshold for using disproportionate military force, should be avoided. As the laws of armed combat do not appear to aid in this matter, I would argue that there is regulatory disconnection.

Another practical issue presented by LAWS directly concerns the principles of discrimination and proportionality: LAWS must unequivocally direct military attacks solely against military objectives and assess the likely collateral harm from an attack vis-à-vis the advantage thereby gained.³⁴ The discrimination concern is the same as that faced by human soldiers: in marginal cases, the demand to distinguish civilians from combatants is more onerous. It would be especially challenging for LAWS to do this when pitted against plainclothes insurgents, particularly in crowded urban environments. Moreover, it is hard to conceive of LAWS making case-by-case judgments as to the proportionality of a particular attack given that military advantage is contextual;³⁵ is it proportionate, for instance, to engage a tank with children riding on it, or an insurgent using civilians as human shields?³⁶ These are confusing and contentious dilemmas for the laws of armed conflict as they apply to soldiers; one cannot devise a simple calculus for the subjective question of how many civilian casualties justify a given military objective.³⁷ Even if, as some argue, LAWS might perform better at objective situational analysis and rapid decision-making,³⁸ it would be wishful thinking to expect even the most optimistic artificial intelligence to answer. This does not make LAWS unlawful *per se*; however, the lack of clarity feeds back into the problem mentioned earlier; if the discrimination and proportionality principles cannot provide an answer, they are little aid in regulation.

The third problem, accountability, is conceptual: people should be held responsible for their wartime acts,³⁹ to avoid exterminating the enemy and to vindicate their grieving relatives.⁴⁰

³⁴ Thurner, above n 3, at 220.

³⁵ *Ibid.*

³⁶ Singer, above n 5, at 389 and 402.

³⁷ Thurner, above n 3, at 223.

³⁸ Galliot, above n 6, at 153-154.

³⁹ Walzer, above n 7, at 287-288.

⁴⁰ Robert Sparrow "Killer Robots" (2007) 4(1) *Journal of Applied Philosophy* at 67.

Failure to identify those responsible for war crimes renders prosecution moot, and is seen as a failure to live up to the principles of just war; as such, technologies which typically achieve such a failure are, in Sparrow's view, unethical.⁴¹ This complements the concern of the previous paragraph: even if LAWS were forced into morally dubious actions, the cost is alleviated somewhat by the ability to attribute responsibility for any resulting war crimes; conversely, if the latter is impossible, then the moral propriety of using LAWS is greatly exacerbated.⁴² So, do LAWS obscure the responsible party? Sparrow argues that they do: if an autonomous machine erroneously delivered misleading information, or independently decided to commit a war crime, holding human beings accountable would be conceptually problematic. Firstly, it would be difficult to hold the programmers responsible: the possibility of a machine attacking the wrong targets might be acknowledged and liability limited thereby; additionally, the notion of an autonomous system suggests that the choices of the machine could constitute a *novus actus*, as the programmers might have no control over such choices.⁴³ Alternatively, the officer ordering the machine's deployment could be held strictly liable for any wrongdoing, as is presently the case for artillery shells which land off target.⁴⁴ However, Sparrow argues that '*the autonomy of a machine implies that its orders do not determine (although they obviously influence) its actions*', with greater autonomy increasing the risk that its commander be held liable for actions of machines over which they had no control.⁴⁵ It may be likewise difficult to hold the commanding officer responsible. The third option is to blame the robot – however, lacking a *mens rea*, this seems legally and morally absurd. Sparrow suggests, therefore, that LAWS create a "responsibility gap" where no blame can be put, implying that they are incompatible with LOAC's notion of responsibility.

4 POTENTIAL SOLUTIONS

⁴¹ Ibid.

⁴² Vincent C. Miller "Autonomous Killer Robots Are Probably Good News" in Ezio Di Nucci and Filippo Santoni de Sio (eds.) *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons* (London: Ashgate, 2016) at 7.

⁴³ At 69-70.

⁴⁴ At 70.

⁴⁵ Ibid.

We have seen how conceptual and practical problems with LAWS which IHL is inept to deal with, or outright incompatible with. What should the regulatory response be, and how might LOAC be changed in order to address the inadequacies above? As we have seen, there are marginal cases wherein the principles of just war seem to reach an impasse; that is, where the balance of benefits and harms is nebulous and subjective. In response to this many have advocated for what Brownsword and Goodwin might call “knee-jerk precautionary response”⁴⁶ – namely, the outright banning of LAWS for the foreseeable future. A notable proponent is the open letter mentioned in the outset, but also Sparrow and a UN Special Rapporteur who recommended in 2013 that states should place national moratoria on LAWS pending superior information and regulation.⁴⁷ This proposition has been met with criticism, firstly on the basis that it precludes the various benefits that LAWS provides (and moreover makes states who conform to such a ban vulnerable to those that do not) and secondly on the grounds that it is unreflective of responses to risk in other areas of civil life. Miller and Galliot each suggest that Sparrow and others overemphasise the conceptual importance of responsibility for wrongdoing in all cases; they cite examples (e.g. civil engineering) where incidences of poor design may nonetheless give rise to no liability due to remoteness of the risk that eventuated. Miller argues that ‘if a technology produces rare cases of killings where no person is responsible, this do [sic] not by itself compel us to ban the use of this technology’.⁴⁸ While ruling out systems which systematically or frequently produce undesirable responsibility gaps, he suggests the ability to acquire and store full data records of LAWS’ action and inputs may actually make allocation of responsibility (and resulting prosecutions) easier where such gaps are absent.⁴⁹ Miller’s preferred regulatory response should be at the international level, incorporating technical standards on the part of manufacturers into IHL to attempt to minimise a machine’s ability to target non-combatants or to report back to a supervising person for directions when confronted with a marginal case;⁵⁰ likewise, national rules of engagement should adopt similar standards to prevent attacks in circumstances where its software could have problems

⁴⁶ Roger Brownsword and Morag Goodwin *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press, Cambridge, 2012) at 47.

⁴⁷ Christof Heyns *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions on lethal autonomous robotics* UN Doc A/HRC/23/47 (9 April 2013) at 118.

⁴⁸ Miller, above n 42, at 9.

⁴⁹ *Ibid.*

⁵⁰ At 10.

distinguishing between military and civilian objects.⁵¹ Manufacturers and operators in breach would be indictable, while incidents falling outside of these would constitute accidents.⁵² Galliot suggests we should go further, and revise the legal notion of responsibility altogether: as ‘relevant programmers, designers, manufacturers and commanding officers are all responsible to some degree or extent’, through their individual relationships to LAWS, he recommends a more complex notion of collective responsibility capable of acknowledging the holistic contributions of human agents, systems and organisations to a particular incident.⁵³ On this basis Galliot suggests that responsibility *can* be apportioned fairly, alleviating Sparrow’s final concern. He does not, however, give much analysis on what his system will look like. Presenting as it does a great step away from conventional views on culpability, it may be a more challenging proposition to implement than Miller’s more orthodox approach, while achieving the same benefit.

The remaining problem, that emotional distance from the battlespace risks lowering operators’ inhibitions to disproportionate force, is admittedly also true for manned weapons that act at a distance.⁵⁴ As the ultimate extension of such weapons, LAWS are of concern; however keeping soldiers out of harm’s way is an obvious concern of LOAC. It would be counterintuitive to claim that the mental and physical suffering caused to combatants in closer proximity to battlespaces is a necessary evil of preventing war crimes. The fear of increasing the likelihood of atrocities, while morally compelling, seems to be out of LOAC’s grasp in this regard, and presents, in my view, the strongest argument for prohibitive regulation. That said, this outcome seems unlikely; the weight accorded to the benefits to personnel means moratoria may not serve states’ interests. If this problem is given any regulatory response I predict it will imitate the oversight currently given to remotely-piloted machines. Beyond codifying such measures my conclusion is that this is a problem where conventional LOAC has little application.

5 CONCLUSION

⁵¹ Thurner, above n 3, at 221.

⁵² Miller, above n 42, at 10.

⁵³ Galliot, above n 6, at 224.

⁵⁴ Miller, above n 42, at 12.

LOAC's current adequacy is mixed. While states can and do practice ongoing legal reviews and reforms of military technology, LAWS poses challenges that strike the heart of *jus in bello*, which obscures any clear solutions. LOAC also seems unequipped to deal with the dehumanising effect of war fought at a distance. Full prohibition of LAWS, however, is disproportionate to these challenges. The most efficacious regulatory responses will likely impose greater responsibility and oversight on both producers and users of LAWS, at both national and international levels.

BIBLIOGRAPHY

STATUTES

Geneva Conventions Act 1958.

TREATIES

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts 1125 UNTS 3 (signed 27 November 1978, entered into force 7 December 1979).

TEXTS

Roger Brownsword and Morag Goodwin *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press, Cambridge, 2012).

Duane Cady *From Warism to Pacifism: A Moral Continuum* (Temple University Press, Philadelphia, 2010).

Jai Galliot *Military Robots: Mapping the Moral Landscape* (Ashgate Publishing Ltd, Surrey, 2015).

Robert McLaughlin and Hitoshi Nasu *New Technologies and the Law of Armed Conflict* (T.M.C. Asser Press, The Hague, 2014).

P. W. Singer *Wired For War: The Robotics Revolution and Conflict in the Twenty-first Century* (The Penguin Books, New York, 2009).

Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 2006).

ARTICLES

Daniel Brunstetter and Megan Brun, “The Implications of Drones on the Just War Tradition”, (2011) *Ethics and International Affairs* 25(3).

Rebecca Crootof "The Killer Robots Are Here: Legal and Policy Implications" (2015) 36 *Cardozo L. Rev.* 1837.

International Committee of the Red Cross Geneva [“IIRC”] “A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977” (2006) 88 *International Review of the Red Cross* 864.

Vincent C. Miller “Autonomous Killer Robots Are Probably Good News” in Ezio Di Nucci and Filippo Santoni de Sio (eds.) *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons* (London: Ashgate, 2016).

Robert Sparrow "Killer Robots" (2007) 4(1) *Journal of Applied Philosophy*.

OTHER

Vincent Boulanin and Maaïke Verbruggen “Article 36 Reviews: Dealing with the challenges posed by emerging technologies” *Stockholm International Peace Research Institute* (December 2017) <www.sipri.org/sites/default/files/2017-12/article_36_report_1712.pdf>.

Autonomous Weapons: An Open Letter From AI & Robotics Researchers” *Future of Life Institute* <<https://futureoflife.org/open-letter-autonomous-weapons/>>.

Christof Heyns *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions on lethal autonomous robotics* UN Doc A/HRC/23/47 (9 April 2013).