Historical and Contemporary Reflections on Lethal Autonomous Weapons Systems

https://www.e-ir.info/2020/04/15/historical-and-contemporary-reflections-on-lethal-autonomous-weapons-systems/

JOHN R. EMERY, APR 15 2020

This article is part of a series.

The issue of the advent of lethal autonomous weapons systems (LAWS) poses significant challenges to foundational assumptions about international humanitarian law, just war principles, and military strategy. Peter Lee et al. in their article, have put forth an important set of LAWS principles formed as part of the Canberra Working Group as a middle-ground and pragmatic approach to regulating the *use* of LAWS. In their: 'Guiding Principles for the Development and Use of Lethal Autonomous Weapon Systems' the authors strive to tackle the *jus in bello* principles of LAWS without taking a position on broader questions of the legitimacy or utilization of LAWS at the *jus ad bellum* stage. In essence, they take the pragmatic assumption that the ethics of deploying LAWS to go to war in the first instance is separate from the ethics that govern their use *in* war. As those that advocate for a preemptive ban on LAWS, like Thompson Chengeta and the Campaign to STOP Killer Robots, this lack of a *jus ad bellum* position is indeed a normative stance, that one cannot sidestep. Although I do believe that a preemptive ban on LAWS is both necessary and ethically preferable, pragmatically it appears as though there is a lack of will by many key states to pursue this path. Thus, I will argue that the Canberra Working Group's 'Guiding Principles' are indeed a step in the right direction, yet they fall short of some empirical obstacles.

In what follows I will demonstrate via concrete historical and contemporary examples of technologies of war, which principles are most promising, and which are most fraught. By exploring the rise of the submarines in the late 19th and early 20th centuries, I argue that LAWS Principles 1 – International Humanitarian Law (IHL) Applies to LAWS – and 6 – Appropriate Use of LAWS is Context Dependent – have a strong historical precedent that leaves me optimistic. In contrast, I will analyze the U.S. utilization of collateral damage estimation algorithms, colloquially known as 'Bugsplat', to problematize principles 2 – Humans are Responsible for the Employment of LAWS – 3 – The Principle of Reasonable Foreseeability Applies to LAWS – and 5 – Command and Control Accountability Applies to LAWS. While these empirical obstacles are no doubt at the forefront of both sides of the LAWS debate it is useful to see how Bugsplat enabled a problematic ethics and provides cautious message for the feasibility of even the Canberra Working Group position on LAWS. Ultimately, I believe that we should to draw upon both the historical precedent, and recent utilization of technologies of war, in order to think through what *ought* to govern their ethical use if a ban cannot be achieved.

The Historical Case of Regulating Submarine Warfare

The history of international humanitarian law and ethics of war is fraught with contestation; yet, I argue that three principles oriented the pre-IHL laws of war were: chivalry, necessity, and humanity. What past and future technological imperatives of war indicate is that the balance between these three ideals shifts with respect to specific technological and cultural possibilities of the particular time and context. The advent of submarine warfare of the late 19th C. and early 20th C. rules of naval warfare is instructive here for dilemmas of AI and LAWS today. Specifically, there was no preemptive ban on submarines, but customary rules grew out of these three principles, which was later somewhat codified in various rounds of international law and the laws of naval warfare in 1899, 1907, 1909 etc. Submarines were an answer to the technological development of hardened ships in the 19th C. In conventional naval

Historical and Contemporary Reflections on Lethal Autonomous Weapons Systems
Written by John R. Emery

warfare, the opposing side was obligated to pick up the survivors of a sunken enemy vessel. However, the technological development of submarines led to an instantaneous evolution in customary international law. Because for submarines, leaving enemy survivors was a violation of the law, but picking them up left the submarines vulnerable to sinking and there was 'no room at the inn' so to speak when it came to submarines.

Thus, the nature of naval warfare changed completely by WWI, whereby the picking up survivors rule still held for the original case of surface ships, but it *could not* hold for submarines. This represented and instantaneous evolution of the customary laws of war based upon a technological imperative. Thus, in the balance of chivalry, necessity, and humanity, technology tipped the scales in favor of necessity at the expense of humanity. Humanity is conditional; the point of IHL is to institutionalize the conditions under which we can hope to be humane in warfare. We want to take humanitarian principles seriously, but many of us tend to be Benthamites (utilitarian's) in war, neglecting the fact that we are craftspeople of international order. IHL is a highly crafted activity that is dependent on science, but it is the craft and technique and skill that we establish the conditions of necessity and humanity today with respect to LAWS.

This historical submarine warfare example demonstrates that technological imperatives in war have a tendency to shift our way of thinking and adapting the laws and ethics of war. In the Canberra document as discussed by Lee et al.'s piece, Principle 1 – IHL applies to LAWS – and Principle 6 – appropriate use of LAWS is context dependent – are upheld by this unique historical precedent, which is a useful lens to examine issues of LAWS today. First and foremost, technology is often first utilized in war and then restrained subsequently, such as the Chemical Weapons Convention. There is an aura surrounding LAWS that once you release killer robots into the wild, there is no reigning them in and I believe this to be an overinflated assertion. Regulating new technologies of warfare is never easy, it takes time, international agreement, institutionalization, and adherence; yet we have done it before without a preemptive ban and can do it again if necessary. According to Principle 1, IHL must apply to LAWS including military necessity, humanity, distinction, and proportionality irrespective of the type of conflict.

Hence, LAWS must remain within compliance of established IHL and customary international law will evolve as LAWS inevitably present new dilemmas that have not been previously accounted for in battle. Secondly, principle 6 is supported; although it did take two World Wars to effectively constrain submarine warfare, subsequent evolutions in rules, norms, and ideas led to a far more restrained weapons system that was able to take greater precautions to minimize incidental loss of life to civilians. Institutionalizing the premise of Principle 6, that LAWS must be context-dependent and not a general program that does not have various levels of command authorization based on the level of risk to civilians, must be upheld. Ultimately, I believe that the rise of submarine warfare presents a unique historical case of technology exploding onto the scene sparking debate and regulation over the coming decades. As LAWS are deployed on the battlefield in new and unpredictable ways, I am optimistic that these Canberra principles 1 and 6 will largely be upheld.

The Contemporary Case of Algorithms of Warfare

Where I believe the principles regulating LAWS ultimately falls short, is its reliance on placing the human at the center of LAWS. While laudable, putting a human in-the-loop or on-the-loop in autonomous systems, it attempts to solve the genuine dilemmas like accountability for killing of innocents, moral agency, and socio-technical interactions. Principle 2 makes the correct judgment (in my view) that LAWS can never be moral *agents* and responsibility cannot be delegated to the technology itself. Placing 'meaningful human control' at the center of many of the debates of AI and LAWS on its surface appears to solve the messy problems that move beyond the scope of our vocabulary of legality and ethics as we struggle to wrap our heads around these dilemmas.

However, there is overwhelming evidence in areas outside the arena of war how humans defer to technology. The *automation bias* means that there is a strong tendency of humans to defer to automated technology; such that humans assume positive design intent even when the tech is malfunctioning, and this holds true even when presented evidence of a system's failure. More troubling, "[a]utomation bias occurs in both naive and expert participants, [it] cannot be prevented by training or instructions, and can affect decision making in individuals as well as in teams." The fact that we treat AI generally as a 'black box' that assumes that machine-driven software is going to offer better judgment than humans. This has major implications for Principle 3, in that we assume the technology

Historical and Contemporary Reflections on Lethal Autonomous Weapons Systems
Written by John R. Emery

will be better than humans neglecting the human assumptions written into the code initially and the ways in which reasonable foreseeability is obfuscated by a sort of faith in technology. This automation bias often degenerates into wishful thinking, or more problematically for LAWS into an opportunistic misuse of technology to validate existing practices.

Finally, Principle 5 (concerning command and control accountability) will often degenerate into an opportunistic misuse of technology. I have conducted some research concerning ethics of due care in war the U.S. utilization of collateral damage estimation algorithms, known colloquially as 'Bugsplat.' What I argue, is that it enables decision-makers to tick the box of ethical due care by deferring to the algorithms. While the Principle 5 is indeed noted as an ideal principle, the problematic ways in which that Bugsplat has been utilized over the past two decades by U.S. military decision-makers should give us caution as to what LAWS enables rather than constrains when it comes to accountability.

Anecdotally, During the initial 'shock and awe' campaign of the 2003 Iraq War, the U.S. military ran the algorithmic collateral damage estimation algorithm *Bugsplat,* which estimated the probable number of civilians that would be killed in a given kinetic strike, it was simply to tick the box of ethical due care. On opening day, the estimations presented to Gen. Tommy Franks indicated that 22 of the 30 projected bombing attacks on Iraq would produce what they defined as heavy Bugsplat – that is, more than 30 civilian deaths per raid. Franks said, "Go ahead, we're doing all 30." There is much more supporting evidence that the use of this algorithmic technology was utilized to defer accountability and opportunistically validate existing practices in troubling ways that is beyond the scope of this discussion. Ultimately, Bugsplat and how it impacted ethical decision-making in war in a negative way (though it is discussed as protecting civilian life) should give us all pause for the prospect of accountability for killing with a future of LAWS.

Conclusion

In the end, there is always uncertainty in warfare. The risk I see for LAWS based on the empirical evidence from Bugsplat, is that technology offers military decision-makers an alluring appeal to technological fixes to ethico-political dilemmas of killing in war. The assumption that LAWS would be at least as good or better than humans at life and death decisions relies on a disconnect between how these algorithms function and how they are utilized. Human judgment is never eliminated from the equation, but it transfers judgment from military practitioners to computer programmers. In my assessment, historically Principles 1 and 6 of the Canberra Working Group are upheld, Principles 2, 3, and 5 are challenged by empirical realities of contemporary algorithmic warfare. While these are ideal-type prescriptions to restrain the utilization of LAWS, which the authors duly acknowledge, we must be aware of the unintended consequences of leaving the human-centered AI as the solution to LAWS dilemmas. While a preemptive ban tackles the *jus ad bellum* normative stance against laws, my pessimistic view is that the Canberra *jus in bello* principles will unfortunately be necessary. In the end, though we take varied stances, the consensus seems to be that LAWS will present new legal and ethical dilemmas of warfare; we all have no idea how precisely that will pan out in the coming years.

About the author:

John R. Emery received his Ph.D. in political science from the University of California, Irvine in 2019. He is currently a Jerome and Hazel Tobis Fellow at UC Irvine's Interdisciplinary Center for the Scientific Study of Ethics and Morality. Dr. Emery's research agenda lies at the intersection of historical International Relations, technologies of war, and international ethics. His research explores how evolving military technologies like AI, Big Data, and lethal autonomous weapons systems are constituted by our epistemological and philosophical assumptions of IR and international law, which his work interrogates. His examination of drones, IR theory, and ethics of war has been published in *Ethics & International Affairs*, *Peace Review*, and in edited volumes published by Georgetown University Press and New York University Press. More information available at: www.EmeryJohnR.com.

Historical and Contemporary Reflections on Lethal Autonomous Weapons Systems
Written by John R. Emery