‘Effective, Deployable, Accountable: Pick Two’: Regulating Lethal Autonomous Weapon Systems

https://www.e-ir.info/2021/08/12/effective-deployable-accountable-pick-two-regulating-lethal-autonomous-weapons-system/

JOHN WILLIAMS, AUG 12 2021

Almost every keen cyclist knows pioneering US engineer Keith Bontrager’s famous observation about bicycles: ‘strong, light, cheap: pick two. If they don’t know it, they’ve experienced its effects at their local bike shop’s checkout when they upgrade any components. The current state of regulatory debate about Lethal Autonomous Weapons Systems (LAWS) looks to be increasingly locked into a similar two-fold choice from three desirable criteria: ‘effective, deployable, accountable: pick two’. However, unlike Bontrager’s bicycles, where the conundrum reflects engineering and material facts, the regulatory debate entrenches social-structural ‘facts’ that make this two-from-three appear inescapable. This article explains how the structure of the LAWS regulatory debate is creating a two-from-three choice, and why the one that holds the most potential for containing the dangers LAWS may create – accountability – looks least likely to prevail. Effective and deployable, just like strong and light amongst cycling enthusiasts, are likely to win out. It won’t just be bank balances that ‘take the hit’ in this case, but, potentially, the bodies of our fellow human beings.

Two key assumptions underpin my claim about an increasingly rigid debate over LAWS regulation. Firstly, LAWS are a realistic prospect for the relatively near-term future. Weapons systems that, once activated, are able to identify, select and engage targets without further human involvement have been around for at least forty years, in the form of systems that target incoming missiles or other ordnance (e.g. Williams 2015, 180). Systems such as Phalanx, C-RAM, Patriot, and Iron Dome are good examples of such systems. These are relatively uncontroversial because their programming operates within strictly defined parameters, which the systems themselves cannot change, and targeting ordnance typically raises few legal and ethical issues (for critical discussion see Bode and Watts 2021, 27-8). LAWS, as I am discussing them here, move outside this framework. Existing and foreseeable AI capabilities, eventually including techniques such as machine learning through deep neural networks, mean LAWS may make decisions within far more complex operational environments, learn from those decisions and their consequences, and, potentially, adjust their coding to ‘improve’ future performance (e.g. Sparrow 2007; Human Rights Watch 2012, 6-20; Roff 2016). Those sorts of capabilities, combined with advanced robotics and state-of-the-art weapons systems point towards LAWS not just to defend against incoming ordnance, but, frequently in conjunction with human combatants, to engage in complex operations including lethal targeting of humans. That targeting may include LAWS that directly apply kinetic effect against their targets – the ‘killer robots’ of sci-fi and popular imagination – but can also extend to include systems where AI and robotic functions provide mission-critical and integrated support functions in systems and ‘systems of systems where a human-operated weapon is a final element.

Secondly, I assume efforts to ban the development and deployment of LAWS will fail. Despite a large coalition of NGOs, academics, policymakers, scientists, and others (e.g. ICRAC, iPRAW, Future of Life Institute 2015) LAWS development is more likely than not. Amandeep Singh Gill (2019, 175), former Indian Ambassador to the UN Conference on Disarmament and former Chair of the Group of Governmental Experts (GGE) on LAWS at the UN Convention on Certain Conventional Weapons (CCW), stresses how:

The economic, political and security drivers for mainstreaming this suite of technologies [AI] into security functions are simply too powerful to be rolled back. There will be plenty of persuasive national security applications – minimizing casualties and collateral damage …, defeating terrorist threats, saving on defense spending, and

protecting soldiers and their bases – to provide counterarguments against concerns about runaway robots or accidental wars caused by machine error.

Appeals to the inherent immorality of allowing computers to make life and death decisions about human beings, often framed in terms of human dignity (e.g. Horowitz 2016; Heyns 2017; Rosert and Sauer 2019), will fall in the face of ostensibly unstoppable forces across multiple sectors making incorporating AI into ever more aspect of our daily lives almost inevitable. From 'surveillance capitalism' (Zuboff 2019) to LAWS, human beings are struggling to find ways to effectively halt, or even dramatically slow, AI's march (e.g. Rosert and Sauer 2021).

Effective

LAWS' potential military effectiveness manifests at strategic, operational, and tactical levels. Operating at 'machine speed' means potentially outpacing adversaries and acquiring crucial advantages, it enables far faster processing of huge quantities of data to generate new insights and spot opportunities, and it means concentrating military effect with greater pace and accuracy (e.g. Altmann and Sauer 2017; Horowitz 2019; Jensen et al 2020). Shifts, even temporary, in delicate strategic balances between rival powers may appear as unacceptable risks, meaning that for as long as adversaries are interested in and pursuing this technology, their peer-rivals will feel compelled to do so too (e.g. Maas 2019, 141-43). Altmann and Sauer (2017, 124) note, 'operational speed will reign supreme'. The 'security dilemma' looms large, reinforcing amongst leading states the sense they dare not risk being left behind in the competition to research and develop LAWS (e.g. Altmann and Sauer 2017; Scharre 2021). Morgan *et al* (2020, xvi) argue the US, for example, has no choice but to, '… stay at the forefront of military AI capability. … [N]ot to compete in an area where adversaries are developing dangerous capabilities is to cede the field. That would be unacceptable'. Things likely look the same in Moscow and Beijing. Add concerns about potential proliferation to non-state actors (e.g. Dunn 2015), and the security dilemma's powerful logic appears inescapable.

Of course, other weapons technologies inspired similar proliferation, strategic destabilization, and conflict escalation concerns. Arms control – a key focus for current regulatory debate – has slowed the spread of nuclear weapons, banned chemical and biological weapons, and prohibited blinding laser weapons before they were ever deployed (e.g. Baker et al 2020). International regulation can alter the strategic calculus about what weapons do and do not appear effective and persuade actors to deny themselves the systems in the first place, or limit their acquisition and deployment, or give them up as part of a wider deal that offers a better route to strategic stability. LAWS present specific arms control challenges because they incorporate AI and robotics technologies offering many non-military opportunities and advantages that human societies will want to pursue, potentially bringing major benefits in addressing challenges in diverse fields. Key breakthroughs are at least as likely to come from civilian research and development projects as from principally military ones. That makes definitions, monitoring, and verification harder. That is not a reason not to try, of course, but it does mean effective LAWS may take many forms, incorporate inherently hard to restrict technologies, and offer possibly irresistible benefits in what the security dilemma presents as an inescapably competitive, militarized, and uncertain international environment (e.g. Sparrow 2009; Altmann 2013; Williams 2015; Garcia 2018; Gill 2019).

Combining with the idea of the inescapable security dilemma are ideas about the unchanging 'nature' of warfare. Rooted in near-caricatured Clausewitzian thought, war's unchanging nature is the application of force to compel an opponent to do our will and in pursuit of political goals to which war contributes as the continuation of policy by other means (Jensen et al 2020). To reject, challenge, or misunderstand this, in some eyes, calls into question the credibility of any critic of military technological development (e.g. Lushenko 2020, 78-9). War's 'character', however, may transform, including through technological innovation, summarised in the idea of 'revolutions in military affairs'. In this framing, LAWS represent the latest and next steps in a computer-based RMA that can trace its origins to the Vietnam War, and which war's nature makes impossible to stop, let alone reverse. The effectiveness of LAWS is therefore judged in part against a second fixed and immutable reference point – the nature of war – that means technological innovations changing war's character must be pursued. Failing to recognise such changes risks the age-old fate of those who took on up-to-date military powers with outmoded principles, technologies, or tactics.

Deployable

Deployable systems face the challenge of operating alongside human military personnel and within complex military structures and processes where human involvement looks set to continue well beyond plausibly foreseeable technological developments. AI already plays support roles in the complex systems behind familiar remotely piloted aerial systems (RPAS, or 'drones') frequently used for targeted killing and close air support operations such as Reaper. This is principally in the bulk analysis of massive quantities of intelligence data collected by these, and other Intelligence, Surveillance and Reconnaissance (ISR) platforms and through other intelligence gathering techniques, such as data and communications intercepts.

Envisaged deployable systems offering meaningful tactical advantages could take several forms. Increasingly AI-enabled and sophisticated versions of current unmanned aerial systems (UAS) providing close air support for deployed ground forces, or surveillance and strike functions in counter-terrorism and counter-insurgency operations are one example. That could extend into air combat roles. Ground and sea-based versions of these sorts of platforms exist to some extent and the same kind of advantages appeal in those environments, such as persistent presence, long duration, speed of operation, and the potential to deploy into environments too dangerous for human personnel. More radical, and further into the future, are 'swarming' drones utilizing 'hive' AI distributed across hundreds or possibly thousands of small, individually dispensable units that disperse and then concentrate at critical moments to swamp defences and destroy targets (e.g. Sanders 2017). Operating in distinct spaces from human forces (except for those they are unleashed against), such swarms could create chances for novel military tactics impossible when having to deploy human beings, placing human-only armed forces at critical disadvantages. Those sorts of systems potentially transform tactical innovation and operational speed into strategic advantage.

Safely deploying LAWS alongside human combatants presents serious trust challenges. Training and other procedures to integrate AI into combat roles will have to be carefully designed and thoroughly tested if humans are to trust LAWS (Roff and Danks 2018). New mechanisms must ensure human combatants are appropriately sceptical of LAWS' decisions, backed by the capability to intervene to override, re-direct, or shutdown LAWS operating irrationality or dangerously. Bode and Watts (2021) highlight challenges this creates even for extant systems, such as Close-in Weapons Systems and Air Defence Systems, where human operators typically lack key knowledge and understanding of systems' design and operational parameters to exercise appropriate scepticism in the face of seemingly counterproductive or counter-factual actions and recommendations. As systems gain AI power that gap likely widens.

Deployable systems that can work alongside human combatants to enhance their lethal application of kinetic force, in environments where humans are present, and where principles of discrimination and proportionality apply present major challenges. Such systems will need to square the circle of offering the tactical and operational advantages LAWS promise whilst being sufficiently comprehensible to humans that they can interact with them effectively, to build relationships of trust. That suggests systems with specific, limited roles and carefully defined functionality. That may make such systems cheaper and faster to make, more easily maintained, with adaptations, upgrades, and replacements more straightforward. There could be little need to keep expensive, ageing platforms serviceable and up-to-date, as we see with current manned aircraft, for example, where 30+ year service lives are now common, with some airframes still flying more than fifty years after entering service. You also don't need to pay LAWS a pension. This could make LAWS more appealing and accessible to smaller state powers and non-state actors, driving proliferation concerns (e.g. Dunn 2015).

This account of deployable systems, however, reiterates the complexity of conceptualising LAWS: when does autonomous AI functionality turn the whole system into a LAWS? AI-human interfaces may develop to the point where 'Centaur' warfare (e.g. Roff and Danks 2018, 8), with humans and LAWS operating in close coordination alongside one another, or 'posthuman' or 'cyborg' systems directly embedding AI functionality into humans (e.g. Jones 2018) become possible. Then the common assumption in legal regulatory debates that LAWS will be distinct from humans (e.g. Liu 2019, 104) will blur further or disappear entirely. Deployable LAWS functioning in Centaur-like symbiosis with human team members or cyborg-like systems could be highly effective, but they further complicate an already challenging accountability puzzle.

Accountable

Presently deployed systems (albeit in 'back office' or very specific roles), and near-future systems reinforce claims to operational and tactical speed advantages. However, prosecuting and punishing machines that go wrong and commit crimes makes little, if any, sense (e.g. Sparrow 2007, 71-3). Where, amongst humans, accountability lies and how it is enforced is contentious. Accountability debates have increasingly focused on retaining 'meaningful human control' (MHC) (Various formulations of 'X Human Y' exist in this debate, but are all sufficiently similar to be treated together here. See Morgan et al 2020, 43 and McDougall 2019, 62-3 for details). Ideally, accountability should both ensure systems are as safe for humans as possible (those they are used against, as well as those they operate alongside or defend), and enable misuse and the inevitable errors that come with using complex technologies to be meaningfully addressed. Bode and Watts (2021) contest the extent to which MHC exists in relation to current, very specific, LAWS, and are consequently sceptical that the concept can meet the challenges of future LAWS developments.

The idea of an 'accountability gap' is widely discussed (e.g. Sparrow 2007; Human Rights Watch 2012, 42-6; Human Rights Watch 2015; Heyns 2017; Robillard 2018; McDougall 2019). The gap ostensibly arises because of doubts over whether humans can be held reasonably and realistically accountable for the actions of LAWS, when those actions breach relevant legal or ethical codes. MHC is a way to close any accountability gap, and takes many potential forms. The most commonly discussed are:

- Direct human authorisation for using force against humans ('in the loop' control).
- Active, real-time human monitoring of systems with the ability to intervene in case of malfunction or behaviour that departs from human-defined standards ('on the loop' monitoring).
- Command responsibility such that those authorising LAWS' deployments are accountable for whatever they do, potentially to a standard of strict liability.
- Weapon development, review and testing processes such that design failures or software faults could provide a basis for human accountability, in this case extending to engineers and manufacturers.

International Humanitarian Law (IHL) is central to most academic analysis, policy debates and regulatory proposals in the CCW GGE, which has discussed this over a number of years (e.g. Canberra Working Group 2020). However, novel legal means, such as 'war torts' (Crootof 2016) whereby civil litigation could be brought against humans or corporate bodies for the damages arising from LAWS failures and errors also appear in debate.

Whilst some state delegations to the CCW GGE, such as the UK, argue current IHL is adequate to deal with LAWS, a significant minority have pushed for a ban on LAWS, citing the inadequacy of current legal regulation and the risks of destabilisation. The most common position favours close monitoring of LAWS developments or, potentially, a moratorium. Any future systems must meet existing IHL obligations and be capable of discriminate and proportionate the use of force (for a summary of state positions see Human Rights Watch 2020). In parallel, new legal and treaty-based regulatory structures, with IHL as the critical reference point to ensure human accountability, should be developed (GGE Chairperson's Summary 2020). That policy stance implicitly accepts the accountability gap exists and must be filled if LAWS are to be a legitimate component of future arsenals (for details of state positions at the CCW GGE see Human Rights Watch 2020).

Two-From-Three

This picture of effective and deployable systems highlights their compatibility and reflects the position found across a broad spectrum of accounts of the military and security literature on LAWS. Accountability turns this into a Bontragerian two-from-three.

Deployable and accountable LAWS would likely be ineffective. Retaining 'in the loop' control as the surest way of enabling accountability precludes systems offering the transformation to 'machine speed'. 'On the loop' monitoring allows more leeway for speed, but if that monitoring is to retain MHC via human interventions to stop malfunctioning or misbehaving systems before they do serious harm, it only loosens the reins a little. The other options all create post facto accountability for harm that has already occurred, rather than stopping it from happening in the first place, so are inherently second best. All look likely to lead to complex, long-running processes to assess the location,

extent, and nature of responsibility and then to apportion appropriate blame and dispense punishment and/or award compensation to humans already substantially harmed. Years of investigation, litigation, appeals, and political and institutional foot-dragging seem highly likely outcomes. Accountability delayed is accountability denied.

Effective and accountable LAWS would be undeployable. Squaring the circle of machine speed effectiveness with human speed accountability (in whatever form that takes) appears daunting at best, impossible at worst (e.g. Sparrow 2007, 68-9), resulting in LAWS of such byzantine complexity or so compromised in functionality as to make them largely pointless additions to any military arsenal. Taking advantage of the strategic, operational, and tactical opportunities of LAWS looks likely to necessitate accepting a very greatly reduced level of accountability.

Conclusion

So, which two to pick? The best answer here may be to return to the idea that, unlike making bicycles, this two-from-three challenge is not constrained by the brute facts of physical materials and engineering processes. The arguments for effective and deployable systems appeal to material-like arguments via the ostensibly inescapable structural pressures of the security dilemma and the military necessity for maximising speed in the exploitation of operational and tactical advantage given war's immutable 'nature' but changing 'character'. Adversaries, especially those less likely to be concerned about accountability in the first place (e.g. Dunn 2015; Harari 2018; Morgan et al 2020, xiv, xv, xvii, 27) may gain more effectiveness from more deployable systems. The supposedly inescapable security dilemma and speed-based logics of war bite again.

LAWS regulation looks, at present, as though it may be an object lesson in the risks of seeing ideational social-structural phenomena as material and immutable. Escaping 'effective, deployable, accountable: pick two', requires a major change in the perspectives on the nature of the international system and war's place within it amongst political and military leaders, especially those in states such as the US, Russia, and China at the forefront of LAWS research and development. There seems a very limited reason for optimism about that, meaning that the regulatory challenge of LAWS looks, at best, to be about harm reduction from the development and deployment of LAWS through creating incentives to try and establish a culture of IHL compliance in design and development of LAWS (e.g. Scharre 2021). More far-reaching and radical change to the LAWS debate potentially involves some quite fundamental re-thinking of the nature of the debate, the reference points used (e.g. Williams 2021), and, first and foremost, a willingness to break free from the ostensibly material and hence inescapable pressures of the nature of war and the security dilemma.

References

Altmann, J. (2013). "Arms Control for Armed Uninhabited Vehicles: an Ethical Issue." Ethics and Information Technology 15(2): 137-152.

Altmann, J. and F. Sauer (2017). "Autonomous Weapon Systems and Strategic Stability." Survival 59(5): 117-142.

Baker, D.-P., et al. (2020). "Introducing Guiding Principles for the Development and Use of Lethal Autonomous Weapons Systems." E-IR https://www.e-ir.info/2020/04/15/introducing-guiding-principles-for-the-development-and-use-of-lethal-autonomous-weapon-systems/.

Bode, I. and T. Watts (2021). Meaning-*less* Human Control: Lessons from Air-Defence Systems on Meaningful Human Control for the debate on AWS. Odense, Denmark, University of Southern Denmark in collaboration with Drone Wars**:** 1-69.

Canberra Working Group (2020). "Guiding Principles for the Development and Use of LAWS: Version 1.0." E-IR https://www.e-ir.info/2020/04/15/guiding-principles-for-the-development-and-use-of-laws-version-1-0/.

Dunn, D. H. (2013). "Drones: Disembodied Aerial Warfare and the Unarticulated Threat." International Affairs **89**(5): 1237-1246.

Crootof, R. (2016). "War Torts: Accountability for Autonomous Weapons." University of Pennsylvania Law Review 164: 1347-1402.

Future of Life Institute (2015). Autonomous Weapons: an Open Letter from AI and Robotics Researchers, Future of Life Institute. https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1

Garcia, D. (2018). "Lethal Artificial Intelligence and Change: The Future of International Peace and Security." International Studies Review 20(2): 334-341.

Gill, A. S. (2019). "Artificial Intelligence and International Security: The Long View." Ethics & International Affairs 33(2): 169-179.

GGE Chairperson's Summary (2021). Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System. United Nations Convention on Certain Conventional Weapons, Geneva. Document no. CCW/GGE.1/2020/WP.7. https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2020/gge/documents/chair-summary.pdf

Harari, Y. N. (2018). Why Technology Favors Tyranny. The Atlantic. October 2018.

Heyns, C. (2017). "Autonomous Weapons in Armed Conflict and the Right to a Dignified Life: an African Perspective." South African Journal on Human Rights 33(1): 46-71.

Horowitz, M. C. (2016). "The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons." Daedalus 145(4): 25-36.

Horowitz, M. C. (2019). "When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability." Journal of Strategic Studies 42(6): 764-788.

Human Rights Watch (2012). Losing Humanity: The Case Against Killer Robots. Washington, DC.

Human Rights Watch (2015). Mind the Gap: the Lack of Accountability for Killer Robots. Washington, DC.

Human Rights Watch (2020). New Weapons, Proven Precedent: Elements of and Models for a Treaty on Killer Robots. Washington, DC.

Jensen, B. M., et al. (2020). "Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence." International Studies Review 22(3): 526-550.

Jones, E. (2018). "A Posthuman-Xenofeminist Analysis of the Discourse on Autonomous Weapons Systems and Other Killing Machines." Australian Feminist Law Journal 44(1): 93-118.

Liu, H.-Y. (2019). "From the Autonomy Framework Towards Networks and Systems Approaches for 'Autonomous' Weapons Systems." Journal of International Humanitarian Legal Studies **10**(1): 89-110.

Lushenko, P. (2020). "Asymmetric Killing: Risk Avoidance, Just War, and the Warrior Ethos." Journal of Military Ethics 19(1): 77-81.

Maas, M. M. (2019). "Innovation-Proof Global Governance for Military Artificial Intelligence?: How I Learned to Stop Worrying, and Love the Bot." Journal of International Humanitarian Legal Studies 10(1): 129-157.

McDougall, C. (2019). "Autonomous Weapons Systems and Accountability: Putting the Cart Before the Horse." Melbourne Journal of International Law 20(1): 58-87.

Morgan, F. E., et al. (2020). Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World, RAND Corporation.

Robillard, M. (2018). "No Such Thing as Killer Robots." Journal of Applied Philosophy **35**(4): 705-717.

Roff, H. (2016). "To Ban or Regulate Autonomous Weapons." Bulletin of the Atomic Scientists **72**(2): 122-124.

Roff, H. M. and D. Danks (2018). ""Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems." Journal of Military Ethics 17(1): 2-20.

Rosert, E. and F. Sauer (2019). "Prohibiting Autonomous Weapons: Put Human Dignity First." Global Policy **10**(3): 370-375.

Rosert, E. and F. Sauer (2021). "How (Not) to Stop the Killer Robots: A Comparative Analysis of

Sanders, A. W. (2017). Drone Swarms. Fort Leavenworth, Kansas, School of Advanced Military Studies, United States Army Command General Staff College.

Scharre, P. (2021). "Debunking the AI Arms Race Theory." Texas National Security Review 4.

Sparrow, R. (2007). "Killer Robots." Journal of Applied Philosophy 24(1): 62-77.

Sparrow, R. (2009). "Predators or Plowshares? Arms Control of Robotic Weapons." IEEE Technology and Society Magazine 28(1): 25-29.

Williams, J. (2015). "Democracy and Regulating Autonomous Weapons: Biting the Bullet while Missing the Point?" Global Policy 6(3): 179-189.

Williams, J. (2021). "Locating LAWS: Lethal Autonomous Weapons, Epistemic Space, and "Meaningful Human" Control." Journal of Global Security Studies. On-line first publication at https://academic.oup.com/jogss/advance-article-abstract/doi/10.1093/jogss/ogab015/6308544?redirectedFrom=fulltext

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future and the New Frontier of Power. London, Profile Books.

About the author:

John Williams is Professor of International Relations in the School of Government and International Affairs at Durham University, UK. His research addresses ethical, regulatory, and security issues associated with new and emerging military technologies and contemporary debates within English School theory. His most recent work on LAWS is published in the Journal of Global Security Studies. He has published with Oxford University Press and in the European Journal of International Relations, Review of International Studies, Ethics & International Affairs, the Chinese Journal of International Politics, and Global Policy.