10

Autonomous weapons systems: managing the inevitability of 'taking the man out of the loop'

GEOFFREY S. CORN

Introduction

War and technological development have been indelibly linked for centuries.[1] Military leaders will constantly seek both the means (weapons) and the methods (tactics) of warfare to maximize their full-spectrum dominance over their adversaries. When assessing the value of emerging weapons technologies, mitigating risk to friendly forces has always been perceived as a key benefit. This aspect of weapons technology is increasingly valued in an era of all volunteer forces and a general perception among strategic decision makers that the public is generally averse to friendly casualties,[2] even when force is employed to achieve vital national or international objectives. At the operational level, all commanders seek to husband resources while achieving precision effects,[3] and, therefore, technologies that facilitate producing such effects with limited risk to friendly forces will be highly coveted.

Willingness to accept mortal risk in pursuit of important objectives is, of course, a core ethos of a professional military. One of the greatest burdens of military command is the authority and responsibility to send subordinates into harm's way to achieve such goals, knowing full well that many may lose their lives or be seriously injured while obeying these

[1] See Levin Institute, *Modern Warfare: Globalization 101*, available at www.globalization101.org/modern-warfare/.

[2] See Lieutenant Colonel R. A. Lacquement Jr, *The Casualty-Aversion Myth: US Army Professional Writing Collection*, available at www.usnwc.edu/getattachment/82192134-8122-404a-a139-2fcc2de2fe38/Casualty-Aversion-Myth,-The–Lacquement,-Richard-.

[3] See Joint Chiefs of Staff, *Joint Publication 3–0: Joint Operations I-13-14*, 11 August 2011, available at www.dtic.mil/doctrine/new_pubs/jp3_0.pdf (stating that a commander's job at the operational level is to design, plan and execute all details of the operation).

orders.[4] Issuing such orders and subjecting subordinates to mortal risk is, however, a key aspect of military command. But when technology can empower a commander to accomplish tactical and operational objectives with little or no risk to friendly forces, it should be self-evident why commanders at all levels covet such options.

However, there have always been inherent limits on the extent to which technology may be used as an effective substitute for human action. To date, these limits have generally focused on the ability to control the effects of a weapon system once employed. Thus, the law prohibits use of such weapons as chemicals and other poison gas, air-delivered incendiaries in populated areas and any other weapons that cannot be directed with any reasonable certainty to strike an intended target.[5] Autonomous weapons – weapons with the capacity to utilize

[4] 'It doesn't take a hero to order men into battle. It takes a hero to be one of those men who goes into battle' – General Norman Schwarzkopf.

[5] See Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction, 13 January 1993, 1974 UNTS 317; Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons, 10 October 1980, 1342 UNTS 137, Article 2: 'It is prohibited in all circumstances to make any military objective located within a concentration of civilians the object of attack by air-delivered incendiary weapons'; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Additional Protocol I), 8 June 1977, 1125 UNTS 3, Article 51; International Committee of the Red Cross (ICRC), *Commentary to the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (*AP I Commentary*) (Geneva: ICRC, 1987) 613–28. Additional Protocol I to the Geneva Conventions defines indiscriminate attacks:

> 4. Indiscriminate attacks are prohibited. Indiscriminate attacks are:
> (a) those which are not directed at a specific military objective;
> (b) those which employ a method or means of combat which cannot be directed at a specific military objective; or
> (c) those which employ a method or means of combat the effects of which cannot be limited as required by this Protocol; and consequently, in each such case, are of a nature to strike military objectives and civilians or civilian objects without distinction.
> 5. Among others, the following types of attacks are to be considered as indiscriminate:
> (a) an attack by bombardment by any methods or means which treats as a single military objective a number of clearly separated and distinct military objectives located in a city, town, village or other area containing a similar concentration of civilians or civilian objects; and
> (b) an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated. (Additional Protocol I, Article 51)

artificial intelligence to replicate human cognitive reasoning[6] – present an entirely new dilemma to the regulation of armed conflict.

The rapid advancement in the technologies enabling the development and fielding of such weapons is apparent, and humanitarian law experts are increasingly focused on the capacity of artificial intelligence to replicate human judgment as the most significant concern associated with the development of such weapons.[7] Certainly, this is a critical aspect of assessing the propriety of using technology in warfare. However, from an operational perspective, an arguably more significant consideration is the capacity of the technological substitute to produce a desired tactical effect in a manner that replicates the level of legal compliance expected from the human actor – the solider.[8] In the context of employing lethal (or even less than lethal) combat power in any situation involving risk to civilians and/or civilian property, the concept of legal compliance involves executing combat operations consistent with the principles of distinction, proportionality and precautionary measures under the fundamental law of armed conflict (LOAC). These principles are further implemented by codified and customary LOAC rules, such as the rule of military objective, the prohibition against indiscriminate attacks and the requirement to consider specific precautions prior to launching an attack that places civilians at risk.

Professional armed forces prepare soldiers to comply with these obligations through training and rely on responsible command to create a high probability of such compliance during mission execution.[9] But no amount of training or supervision can eliminate a very basic reality of

[6] See K. Anderson and M. C. Waxman, 'Law and ethics for autonomous weapons systems: why a ban won't work and how the laws of war can', Jean Perkins Task Force on National Security and Law Essay Series, Research Paper no. 2013–11 (2013), 1, available at http://media.hoover.org/sites/default/files/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf.

[7] See *ibid.*, 3–4; See also ICRC, 'Autonomous weapons: what role for humans?', 12 May 2014, www.icrc.org/eng/resources/documents/news-release/2014/05-12-autonomous-weapons-ihl.htm: 'The central issue is the potential absence of human control over the critical functions of identifying and attacking targets, including human targets. There is a sense of deep discomfort with the idea of allowing machines to make life-and-death decisions on the battlefield with little or no human involvement' (internal quotation marks omitted).

[8] The term 'soldier' is used throughout this chapter generically to indicate any belligerent operative, no matter the operatives' military branch of service or whether the operative is subordinate to state or non-state authority.

[9] See Major D. I. Grimes *et al.*, *Law of War Handbook* (Charlottesville, VA: International and Operational Law Department, Judge Advocate General's Legal Center and School, 2005), 218–22, available at www.loc.gov/rr/frd/Military_Law/pdf/law-war-handbook-2005.pdf; UK

human operatives: they are, and have always been, 'autonomous' weapons systems, because all soldiers must exercise cognitive reasoning in the execution of their battlefield tasks. Characterizing a soldier as an autonomous weapon system (AWS) may appear inhuman, but it is in fact quite accurate. The soldier, when coupled with the means of warfare entrusted to his control, becomes a combat system. And, as a human, the soldier is obviously capable of exercising autonomous judgment and decision making. Indeed, the efficacy of the soldier as a combat system is linked to that autonomous reasoning capacity. Soldiers are trained extensively to prepare them to respond effectively to a wide variety of combat challenges, but no training can replicate the demands of combat or all of the variables that will arise. Training is therefore used to enable the soldier to develop judgment skills that will maximize the likelihood that his or her autonomous judgment will be exercised in a manner that contributes to the overall tactical, operational and strategic objectives of his or her command.

Thus, it is impossible to have absolute 'compliance confidence' for even this 'weapon system' – the weapon system with the most advanced capacity to engage in cognitive reasoning and apply that reasoning to the decision-making process related to unleashing lethal and destructive combat power. Nonetheless, the legality of employing the human 'autonomous' weapon is beyond question. Why is this so? The answer seems clear: because of the presumption that their human autonomous reasoning will be exercised in accordance with the standards imposed by responsible command, which, in turn, indicate an exercise of autonomous reasoning framed by the obligations imposed by the LOAC.

This consideration – the confidence that autonomous judgment will be exercised consistently with LOAC obligations – seems to explain the demarcation line between fielded versus conceptual technological substitutes for human battlefield action and is therefore a potentially critical consideration when exploring the evolution and legality of AWS. As will be explained in this chapter, this consideration undermines the credibility of demands for adopting a per se prohibition against autonomous weapons. However, it also necessitates a creative regulatory focus that ensures these future weapons are fielded only if and when it is possible to validate their ability to produce a level of LOAC compliance confidence that is analogous to – or perhaps even greater than – that of the human soldier.

Ministry of Defence, *The Manual of the Law of Armed Conflict* (*UK LOAC Manual*) (London: Ministry of Defence, 2004).

The centrality and logic of this 'LOAC compliance confidence' focus is illustrated by considering the existing demarcation line between permissible and impermissible technological substitutes for human target engagement judgment. It also reveals that where the technology has been unable to produce a desired tactical effect analogous to that of a human actor, the utility of such weapons has been questioned not from a humanitarian perspective but, rather, from a military operational perspective. A prime example of this phenomenon is the anti-personnel landmine. By relying on rudimentary technology – weight or movement activation – this weapon is quintessentially autonomous.[10] Once an anti-personnel landmine is emplaced, it operates solely on its own to decide when to produce its deadly effect. However, because it is incapable of distinguishing between friend, foe, or civilian in a manner analogous to a human actor, these weapons are widely condemned as unlawful.[11] Of equal significance for the purposes of this chapter, the tactical benefit produced by anti-personnel landmines is widely perceived by military leaders as insufficient to justify the risk of injury to civilians and friendly forces as the result of the weapons' inability to engage in anything close to the type of cognitive reasoning expected of a soldier.[12]

What then explains the persistence of some states, the United States most notably, in retaining anti-personnel landmines for use as a permissible weapon of war? The answer seems to be the intersection of technology, operational necessity and desired effect. US commitment to the approach of the Convention on Conventional Weapons to regulating anti-personnel landmines[13] indicates a commitment to leverage

[10] See US Department of Army, *Operator's and Unit Maintenance Manual for Land Mines*, Technical Manual 9-1345-203-12, October 1995, 1–4, available at http://mines.duvernois.org/LandMines.pdf (describing the function of a landmine).

[11] See Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, 18 September 1997, 2056 UNTS 211, Article 1, 242; See also D. Crane, 'Smart mines: yep, that's the ticket', *Defense Review* (10 April 2004), available at www.defensereview.com/smart-mines-yep-thats-the-ticket/ (explaining that the Bush administration's reasoning for declining to sign the global treaty banning landmines is the United States' policy of using 'smart' mines that self-destruct within a relatively short time period).

[12] See Human Rights Watch, 'Retired generals renew call for total antipersonnel mine ban', 22 July 1997, www.hrw.org/fr/news/1997/07/20/retired-generals-renew-call-total-antipersonnel-mine-ban.

[13] See Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, 3 May 1996, 2048 UNTS 93; see also US Department of State, US Landmine Policy, available at www.state.gov/t/pm/wra/c11735.htm. Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 1990, 19 UNTS 1823.

technological advances that create a high probability of producing desired tactical effects in a manner that effectively distinguishes lawful from unlawful targets. This technology, coupled with restrictions on where anti-personnel landmines may be utilized and conditions related to such use, sufficiently offsets the risk of unintended effect to justify continued reliance on this weapon system.[14] Furthermore, the fact that the United States has chosen not to join or endorse the absolute ban on anti-personnel landmines indicates the tactical and operational value the United States believes is still inherent in this weapon system, so long as the technology is sufficiently advanced to mitigate the risk of unintended effects.

Another oft-cited example of an existing weapon relying on technology as a substitute for human target engagement judgment is the Phalanx anti-missile system.[15] This system functions autonomously to intercept missiles directed against US warships, relying on artificial intelligence to trigger the engagement process.[16] Like the anti-personnel landmine, the Phalanx lacks any meaningful capacity to exercise cognitive reasoning to distinguish between a lawful object of attack and civilians and/or civilian property, nor to distinguish between enemy assets and friendly assets. Nonetheless, the system was fielded based on an apparent determination that it was capable of producing an important tactical effect with minimal risk of error – not because of technical capacity but, instead, because the context in which it is anticipated to be employed will rarely, if ever, implicate a risk that such judgment is necessary to comply with the LOAC. Since the nature of the target engagement the weapon is designed for involves reaction to a high-speed, low-flying missile at sea, the likelihood of something other than an imminent attack on the ship it protects triggering Phalanx engagement is extremely remote, as is the risk that such engagement will create a serious risk of civilian collateral damage or incidental injury. Accordingly, the system is embraced because it is capable of producing the desired tactical effect with minimal risk of error, implicating LOAC compliance.

This provides an interesting contrast. For the landmine, this inability accounts for its widespread condemnation and the decision by so many

[14] See Major G. S. Musselman, *Law of War Deskbook* (Charlottesville, VA: International and Operational Law Department, Judge Advocate General's Legal Center and School, 2011), 160–2, available at www.loc.gov/rr/frd/Military_Law/pdf/LOW-Deskbook-2011.pdf.

[15] See Anderson and Waxman, 'Law and ethics for autonomous weapons systems', 1.

[16] See Federation of American Scientists, 'MK 15 Phalanx close-in weapons system (CIWS)', available at www.fas.org/man/dod-101/sys/ship/weaps/mk-15.htm.

states to join the absolute ban on use. No analogous consternation has been triggered by the Phalanx or similar autonomous seaborne anti-missile systems. What explains this distinction? Clearly, it is the fact that the conditions of employment for the Phalanx indicate an insignificant risk of erroneous engagement, resulting from the fact that, unlike the anti-personnel landmine, the geographic and tactical contexts for such employment suggest that only genuine threats would present themselves in a manner trigging engagement.

Neither of these weapon systems, however, has the capacity to replicate human cognitive reasoning or exercise the type of judgment required to distinguish between lawful and unlawful objects of attack.[17] With anti-personnel land mines, this largely explains their widespread condemnation; for the Phalanx, it is largely irrelevant as the context of employment mitigates the negative consequences of this incapacity. However, it seems that the proverbial game is on the verge of changing dramatically. As numerous experts have noted, the rapid evolution of artificial intelligence renders these examples simplistic compared to what is on the weapons horizon.[18] What is anticipated, and what causes so much consternation, is the prospect of an AWS designed to exercise a level of cognitive reasoning analogous to – or, in the view of some, superior to – that of humans.[19]

Many artificial intelligence experts question whether the development of truly autonomous weapons – what some have labelled 'killer robots' – is even conceptually feasible. Based on publicly available information, it does not seem that such weapons will be available for military procurement in the foreseeable future. Indeed, this most likely provides some explanation for the current US Department of Defense (DoD) policy on autonomous weapons requiring some human involvement in the targeting loop, a limitation that seems consistent with an expectation that this requirement will align with the inherent limitations on the foreseeably available technology.

However, if the availability of such weapons does move from the realm of speculation to reality, it is very likely that states, and, more specifically, military leaders, will be attracted by the opportunities they offer to maximize the effects of combat power while mitigating risk to friendly forces. This does not mean that all states will embrace this opportunity.

[17] See Additional Protocol I, Articles 50–1; see also G. S. Corn *et al.*, *The Law of Armed Conflict: An Operational Approach* (New York: Aspen, 2012).
[18] See Anderson and Waxman, 'Law and ethics for autonomous weapons systems', 1.
[19] See *ibid.*, 2.

Even assuming a proposed weapons system offers the capacity to substantially contribute to the achievement of targeting priorities, there may be reluctance – especially among military leaders – to field lethal capabilities that substitute artificial intelligence for human judgment. This reluctance might arise from what may be emerging as a somewhat ironic inverse relationship between the capacity of technology to offer such substitutes and the perceived importance of human judgment in relation to the execution of military operations. This potential tension between the theoretical advantages offered by autonomous weapons and the perception that removing human judgment from the target engagement process will be addressed in more detail below.

How military leaders will respond to the opportunities offered by emerging weapons technology is ultimately speculative. Even considering what might be considered negative consequences of increasingly sophisticated automation, it is still not difficult to imagine that military leaders, and, in turn, their nations, will be tempted to embrace these advances. While critics of truly autonomous weapons believe they pose immense risk to humanity,[20] the perception and reaction of military and civilian leaders and experts will undoubtedly influence how their states react to this concern. As a result, interests of humanity will certainly not be the exclusive consideration influencing the reaction to these weapons and perhaps not even the dominant consideration. Instead, the response to the opportunity to develop and field increasingly autonomous weapons capability will most likely focus on a very different question: whether the positive tactical and operational potential of these weapons can be achieved without a serious risk of unintended unlawful consequences.

This question will only be answered with time, as weapons developers ply their wares to military consumers. However, the risk of development to outpace legal compliance validation is genuine and most likely explains policies such as that adopted by the DoD, which establishes strict oversight requirements related to autonomous weapons development and procurement. Discussed in greater detail later in this chapter, such policies demonstrate the importance of ensuring that military goals, technology and legal oversight are each completely engaged in this development process. Such a process will advance the important interest of ensuring that both developers and consumers of weapons technology understand and demand that the ability to discriminate between lawful and unlawful objects of attack, with minimal risk of error, is a central

[20] *Ibid.*

component to the artificial intelligence responsible for any autonomous engagement decision. In fact, if this consideration does drive autonomous weapons development, it is unrealistic to demand or expect their outright prohibition. It is equally unrealistic to expect that international law will in some manner effectively prohibit autonomous weapons development or substantially circumscribe the use of such weapons (for example, by prohibiting use in proximity to civilians or civilian property). The combined effect of market forces (the immense profit potential associated with the development and marketing of highly advanced weapons systems), casualty aversion (the desire of strategic leaders to minimize the risk to friendly forces when conducting military actions against enemy personnel and capabilities) and rapid crisis management (the interest in addressing emerging threats at the nascent stage) make the development, procurement and utilization of autonomous weapons a very likely evolution of modern warfare.

Those who press for international legal prohibitions on the development and fielding of AWS must recognize the potential tactical and operational value inherent in these weapons. Prohibiting the use of autonomous weapons, even if they can function consistently with LOAC principles, will be considered by many as a further step in the already troubling attenuation between conflict regulation and strategic, operational and tactical realities. In this regard, such prohibitions may undermine the credibility of the law itself. Indeed, even the alternative characterization for the LOAC – international humanitarian law – highlights this concern. For several decades, there has been an ongoing debate over the proper characterization for this branch of international conflict regulation law.[21] While this debate is in large measure now stale, it represents a consistent and continuing concern that the very title of this branch of international law suggests a primary humanitarian objective.[22] For proponents of the LOAC characterization, this is a distortion of the historic foundation and purpose of the law. Instead, they believe, the humanitarian component of the law is only one aspect of the regulation of conflict and that, when considered in proper context, the entire

[21] See US Department of Defense (DoD) Directive 2311.01E, 'DoD law of war program', 9 May 2006, available at www.dtic.mil/whs/directives/corres/pdf/231101e.pdf; See also ICRC, *What Is International Humanitarian Law?* (Geneva: ICRC, July 2004): 'International humanitarian law is also known as the law of war or the law of armed conflict.'

[22] See ICRC, *What Is International Humanitarian Law* (explaining that international humanitarian law exists to protect civilians during armed conflict between nations).

body of law should not be characterized in a manner inferring that it is the law's predominant purpose.[23]

However, this debate, stale as it may be, highlights a much more significant underlying concern, which is the risk that the law will evolve in a manner that may not adequately account for the legitimate interests of armed forces.[24] Humanitarian restraint is undoubtedly a noble goal, but this law ultimately must account for the pragmatic interests of those called upon to engage in mortal combat. Armed conflict involves the inevitable reality that belligerents will seek to impose their will on each other through the application of deadly combat power and that they will do so in the most efficient and effective manner. Humanitarian constraints are, as they have always been, essential to mitigate the suffering caused by this contest, but when these constraints are perceived as prohibiting operationally or tactically logical methods or means of warfare, it creates a risk that the profession of arms – the very constituents who must embrace the law – will see it as a fiction at best or, at worst, that they will feign commitment to the law while pursuing *sub rosa* agendas to sidestep obligations.

As I argued in a prior article, this is why those responsible for advancing the law must be vigilant in preserving the historic symmetry between

[23] See Grimes, *Law of War Handbook*, 2–15 (describing the historical development of the law of armed conflict (LOAC) and its functions to protect human dignity and morality amidst the necessity of war in a civilized society).

[24] This is obviously a highly subjective concern. However, there are indications that concerns over civilian protections may be increasingly distorting the balance between military necessity and humanity. Reluctance by significant military powers to commit to treaties banning certain weapons – such as cluster munitions – reflects some level of inconsistency between the perceived legitimacy of the balance of interests reflected in these treaties. International jurisprudence, and scholarly commentary, also influence the evolution of the law and should be considered in this regard. One such illustration was the International Criminal Tribunal for the former Yugoslavia Trial Chamber judgment in *Prosecutor* v. *Gotovina*, case no. IT-06-90-T, Judgment, 15 Apr. 2011. The response from a group of military experts explicitly raised the concern that the evidentiary foundation for condemning General Gotovina risked imposing a standard of conduct that was incompatible with the reality of military operations. Other examples of this trend include the ongoing assertion of a least-harmful means obligation in relation to targeting enemy belligerents and some of the interpretations of the 'warning obligation' asserted in response to the Israeli Operation Protective Edge in Gaza. See 'Soldiers and human rights: lawyers to the right of them, lawyers to the left of them', *The Economist* (9 August 2014), available at www.economist.com/news/britain/21611096-army-increasingly-feels-under-legal-siege-lawyers-right-them-lawyers-left-them?fsrc=rss|btn# (describing an increase in civil litigation in Britain attempting to harmonize the human rights laws with the LOAC).

military logic and humanitarian constraints.[25] This does not mean that the plight of innocent victims of war – whose numbers do not seem to have abated as the nature of means of warfare have increased in sophistication, precision and lethality[26] – should be subordinated to the military necessities. What it does mean, however, is that these necessities must be carefully assessed to determine the demarcation line between actual military logic and demands for overly broad operational freedom of action.

There are probably no aspects of conflict regulation where this need is more acute than in the development of autonomous weapons.[27] International law experts must be involved in this development process, but they must also be cognizant of the enormous appeal such weapons will present to military leaders. However, military leaders and those responsible for procuring and fielding weapons must also recognize the inherent risk associated with pursuing weapons systems that cannot produce both tactical and strategic benefit. History has demonstrated time and again that there is simply no military utility – and, in fact, immense disutility – in any military measure incapable of producing tactical benefit without ensuring LOAC compliance. This idea applies equally to both tactics (methods of warfare) and weapons systems (means of warfare). Accordingly, no matter how appealing the tactical benefit of an autonomous weapon may appear, its value is illusory unless it can effectively distinguish between lawful and unlawful objects of attack. With military leaders increasingly cognizant of the link between legal compliance, perceptions of legitimacy and strategic success,[28] military leaders should embrace strict limitations on the fielding of these future weapons in order to ensure sufficient LOAC compliance capability. Thus, the ultimate thesis of this chapter reflects the inherent balance of the LOAC itself: humanitarian advocates must avoid the temptation of

[25] See generally G. S. Corn, 'Mixing apples and hand grenades: the logical limit of applying human rights norms to armed conflict', *Journal of International Humanitarian Legal Studies*, 1 (2010), 52.

[26] See E. T. Jensen, 'The future of the law of armed conflict: ostriches, butterflies, and nanobots', *Michigan Journal of International Law*, 35 (2014), 253 (analysing how the law of armed conflict will evolve to address the challenges of evolving weapons technology).

[27] See Anderson and Waxman, 'Law and ethics for autonomous weapons systems', 11–12; see also Jensen, 'The future of the law of armed conflict'.

[28] See US Department of Army, TRADOC Pam 525-3-1: 'The US Army operating concept', 19 August 2010, 28, available at www.tradoc.army.mil/tpubs/pams/tp525-3-1.pdf (indicating the importance of civilian and military leaders operating in unity with a clear understanding of the legal limitations of war).

unrealistic autonomous weapons per se prohibitions, and military leaders must avoid the seductive effect of emerging technology.

Rethinking the locus of compliance validation

Demanding integration of technical characteristics that produce a high level of confidence that the weapon will comply with the LOAC is vital, because the very nature of autonomous weapons inverses the normal locus of human involvement in the weapon utilization process. Article 36 of Additional Protocol I[29] establishes an obligation that all new weapons systems be reviewed to ensure LOAC compliance, but it provides very little guidance on the nature of this review.[30] The review must, however, assess the compatibility of the proposed weapon system with LOAC standards when the weapon is used as intended.[31] For most weapons, this is only a preliminary step in ensuring ultimate employment complies with the LOAC. This is because it is the human involvement in that employment – the exercise of human judgment as to when, where and how to employ the weapon – that will often be far more decisive in ensuring such compliance. In contrast, for truly autonomous weapons, the focal point of LOAC compliance will shift to the development/procurement phase. This is because tactical employment will have a significantly reduced influence on LOAC compliance: tactical commanders will employ these weapons pursuant to established use criteria, but beyond that they will have very little influence on how the weapon executes combat operations. It is therefore essential to reconceive the LOAC compliance model when contemplating the procurement, production and employment of AWS. The process of mitigating the risk of failures in this cognitive process must be tailored to address this inversion of the LOAC compliance influence between the procurement and employment phases of utilization. Doing so will focus compliance efforts at the decisive point in this process: the procurement phase.

[29] See Additional Protocol I, Article 36.

[30] See ICRC, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol 1 of 1977* (Geneva: ICRC, 2006), 933, available at www.icrc.org/eng/assets/files/other/irrc_864_icrc_geneva.pdf: 'Article 36 does not specify how a determination of the legality of weapons, means and methods of warfare is to be carried out.'

[31] See, e.g., US Department of Defense (DoD) Directive 3000.09, 'Autonomy in weapon systems', 21 November 2012, available at www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

This reconception should begin by acknowledging an undeniable reality: that soldiers are themselves AWS. The soldier, like the hypothesized and arguably speculative AWS, is capable of exercising cognitive reasoning. This is obviously inherent in any human being. However, the soldier is not an unconstrained 'autonomous actor'; he/she does not exercise judgment with no parameters. Instead, the soldier operates as an agent of responsible command and, in that capacity, must frame his/her decision-making process within the parameters established by superior command. How the soldier is developed and prepared to exercise this inherently autonomous cognitive capacity without becoming an autonomous actor therefore provides a logical template for the 'preparation' of a weapon system with autonomous cognitive capacity. The goal must ultimately be to ensure the autonomous weapon functions in a manner that, like the soldier, is subordinated to the will and parameters imposed by responsible command. The ability to employ combat power consistent with LOAC obligations is inherent in that superior–subordinate relationship.

Preparing the 'weapon' for legally sound employment

Soldiers, or perhaps more accurately the soldier mentality, are in many ways 'produced' through a development process. Much of the focus of initial entry training is to develop in the soldier a sense of discipline that results in a high degree of confidence that the soldier will be capable of adjusting to the demands of military society. First and foremost among those demands is subordination to the orders of superior leaders. This superior–subordinate relationship is the essence of a military organization and involves the willingness to not only obey orders to employ deadly force on command but also to obey orders to refrain from using such force when individual instinct may be dictating the exact opposite response to a given situation.[32]

Preparing soldiers to obey orders that require them to subordinate their personal self-interest for the greater good of the military unit and mission is a primary responsibility of military leadership. Intrinsic to this preparation process is developing soldiers to obey orders that are

[32] See G. Klein *et al.*, *Enhancing Warrior Ethos in Soldier Training: The Teamwork Development Course* (Research Institute for the Behavioral and Social Sciences, US Department of the Army, 2006); G. Riccio *et al.*, *Warrior Ethos: Analysis of the Concept and Initial Development of Applications* (Washington, DC: Research Institute for the Behavioral and Social Sciences, US Department of the Army, 2004).

intended to ensure respect for, and compliance with, the LOAC. Ensuring that a military unit employs collective violence only for lawfully permissible purposes is therefore equally central to the notion of responsible command.[33] It is therefore unsurprising why operating pursuant to 'responsible command' as part of an organization that complies with the laws and customs of war has been the historic *sine qua non* of qualification for lawful belligerent status.[34]

The link between responsible command, LOAC compliance and lawful combatant status reveals a ground truth about the law: only those individuals capable of autonomous reasoning who have been incorporated into an organization capable of managing the exercise of that reasoning should be granted the privilege of engaging in hostilities.[35] It is through this command–subordinate relationship that the law establishes a high degree of confidence that the 'autonomous human' will not use the power entrusted to him/her in a truly unconstrained manner but, rather, will instead exercise that autonomy within the boundaries imposed by superior authority and intended to ensure mission accomplishment within the LOAC legal framework.

The process of influencing and framing the exercise of cognitive reasoning will function much differently for autonomous weapons than for the soldier. As a result, the superior–subordinate relationship will not produce an analogous effect on the autonomous weapon. For the soldier, initial training prior to 'fielding' merely lays the foundation for the ongoing process of framing or shaping the exercise of cognitive reasoning and independent judgment.[36] Unit commanders then build upon this foundation by exercising their responsibility to further develop the soldier through the continuation of the training process. Of equal

[33] See Additional Protocol I, Article 87; see also *In re Yamashita*, 327 US 1, 16–17 (1946); see also ICRC, *AP I Commentary*, 1005–16.

[34] See Additional Protocol I, Article 43; Convention (IV) Respecting the Laws and Customs of War on Land and Its Annex: Regulations Concerning the Laws and Customs of War on Land, 18 October 1907, 32 Stat 1803, Articles 1–3; see also ICRC, *Commentary: Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts* (Geneva: ICRC, 1987); see generally G. S. Corn, 'Unarmed but how dangerous? Civilian augmentees, the law of armed conflict, and the search for a more effective for permissible civilian battlefield functions', *Journal of National Security Law and Policy*, 2 (2008), 257.

[35] See Additional Protocol I, Article 87; see also ICRC, *AP I Commentary*, 1017–23.

[36] See US Department of Army, *Soldier's Manual of Common Tasks STP 21-24-SMCT: Warrior Leader Skills Level 2, 3, and 4*, September 2008, available at www.milsci.ucsb.edu/sites/secure.lsit.ucsb.edu.mili.d7/files/sitefiles/resources/STP%2021-24-SMCT,%20Warrior%20Leader%20Skills,%20Level%202,%203,%20and%204.pdf.

importance, these commanders are expected to establish a command culture that emphasizes not only tactical aggressiveness, but also humanitarian constraint derived from LOAC obligation. The significance of these dual responsibilities is reflected in the doctrine of command responsibility, which holds commanders criminally accountable for subordinate war crimes for failure to effectively execute these two aspects of command responsibility.[37]

Field commanders will not, however, have a meaningful opportunity to influence the exercise of cognitive reasoning and judgment by truly autonomous weapons. Instead, they will merely unleash these weapons when a situation indicates that the purported capability will produce a desired effect.[38] In this regard, the autonomous weapon is perhaps more like the brand new replacement soldier, fresh out of initial entry training, who is fielded and deployed into combat with no opportunity to undergo unit training. In such a situation, the employing commander will rely almost exclusively on the expectation that the initial training effectively prepared the soldier for the complex battlefield judgments he or she will

[37] See Additional Protocol I, Article 87; see also US Department of Army, *Field Manual 27–10: The Law of Land Warfare*, 18 July 1956, para. 2, available at armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm27_10.pdf (the purpose of the law of war 'is inspired by the desire to diminish the evils of war by: (a) Protecting both combatants and noncombatants from unnecessary suffering; (b) Safeguarding certain fundamental human rights of persons who fall into the hands of the enemy, particularly prisoners of war, the wounded and sick, and civilians; and (c) Facilitating the restoration of peace'); *UK LOAC Manual*. See also V. Hansen, 'What's good for the goose is good for the gander – lessons from Abu Ghraib: time for the United States to adopt a standard of command responsibility towards its own', *Gonzaga Law Review*, 42 (2007), 335; Y. Shany and K. R. Michaeli, 'The case against Ariel Sharon: revisiting the doctrine of command responsibility', *New York University Journal of International Law and Policy*, 34 (2002), 797 (summarizing the history of command responsibility).

[38] It is self-evident that a commander employing such a weapon will have the ability to employ the weapon unlawfully. This is inherent in every weapon system and every soldier entrusted to the commander's responsibility. When commanders utilize methods and means of warfare with the intent to violate the LOAC, they will be personally responsible for the ensuing violation under any conception of principal liability. However, methods and means of warfare are normally employed in a manner intended to comply with the LOAC. The much more complex question of command liability (responsibility) arises when the commander does not intend an unlawful outcome, but such an outcome results from subordinate misconduct, which may be the result of employing a lawful weapon system in an unlawful manner. As will be discussed below, liability may also arise when such violations are foreseeable consequences of a leadership failure. This presupposes the commander to whom liability is imputed could have influenced the subordinate in a manner that would have averted the violation. It is the inability to produce this type influence on subordinate reasoning and judgment that distinguishes the truly autonomous weapon from the soldier.

have to make. And even this comparison is somewhat inapposite, as even this brand new member of the unit will be capable of perceiving and internalizing the command culture. Of equal importance, the moment that soldier joins the unit he or she will be subjected to the leadership actions of his superiors, which, even in the shortest period of time, can contribute to shaping the soldier's cognitive reasoning and thereby contribute to either respect for, or aversion to, the LOAC.

Field commanders will have no analogous shaping or influencing opportunity on the artificial intelligence that will dictate the reasoning of an autonomous weapon. Instead, the fielding commander must essentially take on faith the capacity of the weapon to exercise cognitive judgment in a manner that ensures LOAC compliance and, in turn, respect for what is hopefully the command culture. Accordingly, it is the development phase – analogous to the initial training phase for a soldier – that is the decisive point in establishing parameters to ensure cognitive autonomy is exercised within the parameters that ensure LOAC compliance, and in turn both the tactical and strategic interests of the fielding force. As a result, the inputs of military procurement managers, weapons developers and legal advisors must be fully engaged in the weapons development process to ensure the commander employing such a weapon system may do so with genuine confidence that the system will exercise cognitive reasoning consistent with LOAC requirements and at least as effectively as should be expected of the best-trained soldier.

The three LOAC compliance confidence enablers

Accomplishing this goal will require emphasis on three LOAC compliance enablers. First, the military command seeking autonomous weapon capability must carefully and precisely define the tactical effect it seeks to achieve with the weapon, the desired method by which the weapon will produce that effect and any limitations that will be imposed on use of the weapon. In very general terms, this is nothing more than an articulation of the anticipated 'task and purpose' for the weapon, which will in turn define the permissible use of the weapon and, in so doing, facilitate LOAC compliance validation.

This approach is commonly used when issuing tactical orders to subordinate units. One somewhat simplified example is the mantra that the mission of the infantry is to 'close with and destroy' the enemy: the task is to confront the enemy; the purpose is to destroy the enemy. In fact, 'task and purpose' is normally far more refined. For example, for a unit of soldiers, the

commander might indicate that the 'task' is to breach an obstacle, while the 'purpose' is to enable friendly movement to contact with the enemy. Translating this methodology to the autonomous weapon, for example, might result in a 'task' of close with and attack enemy ground forces in a populated area, with the 'purpose' of subduing enemy forces without having to to endanger friendly forces. Other illustrations might include: identify and suppress enemy air defense capabilities; identify and suppress enemy fire support capabilities; acquire and engage with indirect fires enemy command and control in order to disrupt such capability; conduct counter-surveillance acquisition and disable enemy surveillance assets. Although it need not take the precise form of 'task and purpose', maximizing the articulation of the intended tactical function for the AWS will inevitably facilitate the capacity of legal advisors responsible for vetting the weapon system to assess the capacity to comply with the LOAC in the tactical context in which use of the weapon is intended.

This is an important first step in the compliance process, because this articulation will enable those providing operational and legal review and oversight of the development process the ability to assess the inherent risk of LOAC violation associated with the weapon. In other words, this will define the weapon's intended use, which will be the focus of both legal compliance review and employment parameters. This, in turn, will enable the integration of LOAC violation mitigation measures with the weapon system itself or through employment parameters established prior to fielding the weapon.

Such a 'task and purpose' foundation is actually analogous to the preparation of military units for various missions. By training a unit to function in a specific operational and tactical environment, commanders enhance their confidence in legally compliant mission execution.[39] Furthermore, those responsible for task-organizing military units for specific missions are able to enhance the likelihood of legal compliance by refraining, when feasible, from committing a unit trained to participate in one operational context (for example, high-intensity conflict) into a context the unit has not been well prepared for (for example, a peacekeeping mission).[40] Thus, defining the task and purpose of the weapon

[39] See US Dept of Army, *Field Manual 7–0: Training for Full Spectrum Operations* para. 1–30 (12 December 2008) (explaining that the legal complexity of war means that soldiers trained specifically in certain areas, such as stability tasks, will be ineffective when involved in other tasks, such as civil support).

[40] See 'Inquiry into abuse by GI's in Kosovo faults training', *New York Times*, 19 September 2000, www.nytimes.com/2000/09/19/world/19MILI.html.

system will ideally enable the integration of LOAC compliance criteria tailored to that task without undermining the ability to achieve the intended purpose. Or, in the alternative, it will inform field commanders of the situations in which the system may be permissibly used and those where it may not.

Second, LOAC 'compliance standards' must be established at the national level and integrated into the procurement and development process, and it must be clear that no autonomous weapon may be fielded without satisfying these standards. These standards must be imposed to effectuate a level of 'compliance confidence' greater than that of an actual soldier who has passed through the pre-deployment development process. This is because, unlike the soldier, it will be impossible to 'downstream' the fulfillment of this development process to the commander and unit responsible for tactical employment. Integrating compliance mechanisms into the development of the system, and prohibiting the fielding of systems that fail to satisfy these strict requirements, should produce confidence that the system will function consistently with the standards of the 'responsible' tactical or operational command.

This is an apparent focal point within the US DoD as it begins to contemplate the procurement, development and fielding of autonomous systems, as illustrated by the relatively recent DOD Directive 3000.09.[41] This Directive reflects the commitment to establishing LOAC compliance confidence at the pre-fielding phase of employment for any AWS. Most significantly, the Directive requires that any AWS be capable of exercising judgment analogous to that of a human when the consequence of employment implicates LOAC targeting norms, such as distinction and proportionality. Thus, the Directive requires that any future autonomous weapon:

- function pursuant to LOAC requirements, to wit:
  - requirement for proportionality and discrimination
  - can be built into system design or employed only narrowly
- include fail-safes to provide robust protection against possible failures:
  - 100 per cent error free operation not required, but system design must allow for human intervention before unacceptable levels of damage occur
  - robustness harder to ensure than simply ethical decision making?
- is capable of accomplishing military mission.[42]

[41] See DoD, Directive 3000.09. [42] See *ibid.*

How such requirements will be validated, much less developed, is certainly a complex question. But that question is beyond the scope of assessing the potential legality of employing such weapons. Quite simply, if the technology cannot meet these requirements, or there is no way to effectively validate assertions that they have been met, the weapon must not be fielded. Any other decision would be analogous to deploying a soldier so poorly trained that he/she is incapable of navigating the complex environment of battle with any degree of confidence that he/she will do so consistently with the requirements of the mission and the law.

The third critical enabler for ensuring legally compliant use of these weapons is the legal review process. Article 36 of Additional Protocol I has required since 1977 that:

> In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.[43]

This new weapon review requirement is arguably a customary international law obligation[44] and was implemented by the United States (which is not a party to Additional Protocol I) two years prior to 1977.[45] Unfortunately, very few states have implemented a robust program to meet this obligation.[46] One explanation for this gap between obligation and implementation is the reality that most conventional weapons fielded in the three decades following Additional Protocol I coming into force were produced by a relatively small number of nations. Perhaps states procuring such weapons did so on the assumption that if they were produced for the armed forces of the state of production, by implication, they satisfied the legal requirements.

This explanation, of course, is problematic in the abstract, as individual states bear responsibility for independently ensuring newly fielded weapons comply with the LOAC. However, because LOAC compliance

[43] See Additional Protocol I, Article 36.

[44] See ICRC, *Guide to the Legal Review of NewWeapons,* 933; See also J. Henckaerts and L. Doswald-Beck, *Customary International Humanitarian Law* (Geneva: ICRC, 2005), 250 available at www.icrc.org/eng/assets/files/other/irrc_864_icrc_geneva.pdf.

[45] See DoD Directive 3000.09; DoD Directive 5000.01: 'The defense acquisition system', 12 May 2012, available at www.dtic.mil/whs/directives/corres/pdf/500001p.pdf.

[46] See J. McClelland, 'The review of weapons in accordance with Article 36 of Additional Protocol I', *Current Issues and Comments*, 85 (2003), 398, available at www.icrc.org/eng/assets/files/other/irrc_850_mcclelland.pdf.

will be dictated almost exclusively during the weapons development process, the entire legal review process must take on a fundamentally different emphasis in relation to autonomous weapons. Unlike most new weapons systems, autonomous weapons will involve an inherent reduction in the expectation that the tactical employment of the weapon will substantially influence LOAC compliance. Instead, the development phase must be understood as decisive in establishing a high degree of LOAC compliance confidence.

This will present a number of complexities. First among these will be the challenging intersection of law and technology.[47] Lawyers reviewing weapons system legality must understand the nature of the weapon. As artificial intelligence evolves to provide mechanical devices with cognitive capability, it is essential that the technical characteristics of this capability be translated into terms that can be meaningfully understood and critiqued during the legal review process. This is also another aspect of the development and fielding process where the contribution of a clear and carefully defined task and purpose becomes apparent. The legal advisor will ultimately be responsible for advising the procuring commander with an opinion that the capacity of the artificial intelligence satisfies the intended task and purpose with minimal risk of unacceptable error.

This 'LOAC risk' analysis will be central to the credibility of any legal review. Critics of developing and fielding autonomous weapons focus on the inherent risk that these weapons will be incapable of complying with the fundamental LOAC targeting rules, most notably distinction and proportionality.[48] LOAC compliance, however, does not require a zero risk standard. Such a standard would be inconsistent with the requirements related to the most advanced AWS currently fielded: the soldier. Deploying a soldier into hostilities always involves some risk that the soldier will exercise his/her autonomous judgment in a manner that results in an LOAC violation. Demanding that nations guarantee that this will never occur is unrealistic. For the soldier, this risk is mitigated through the training, guidance and the discipline process – the very

[47] See Anderson and Waxman, 'Law and ethics for autonomous weapons systems', 9–11.

[48] See Human Rights Watch, *Losing Humanity: The Case against Killer Robots* (New York: Human Rights Watch, 2012), 30–4, available at www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf; See also J. Foy, 'Autonomous weapons systems: taking the human out of international humanitarian law,' *Dalhousie Journal of Legal Studies*, 23 (2014), 61–5.

essence of leadership. For the autonomous weapon, it must be mitigated through technology validation.

However, I do believe that it is appropriate to demand that the process of reviewing the weapon's capability produces a higher degree of LOAC compliance confidence than that demanded of the soldier. I believe two considerations justify this more demanding standard. First, it is almost impossible to completely assess such risk in humans. While it is undeniable that any human is susceptible to deviations from expected and demanded standards of judgment, this susceptibility is often latent and imperceptible until it unfortunately manifests itself. Examples of latent defects in human judgment producing LOAC violations are unfortunately too common; US Army Staff Sergeant Robert Bales is, perhaps, a prime example of this problem.[49] In contrast, technical aspects of artificial intelligence are susceptible to objective critique and assessment. Information technology experts should be able to 'debug' the artificial intelligence associated with autonomous weapons to a degree of certainty that substantially exceeds that which is possible for human actors. Second, a more demanding standard for LOAC compliance is necessary to provide fielding commanders with a level of confidence that the weapons they employ, but have very little opportunity to influence, will not compromise either their tactical or strategic objectives. Since the fielding commander, unlike the soldier, will have virtually no ability to influence the exercise of autonomous weapons judgment as it relates to LOAC compliance, the degree of compliance confidence prior to ever fielding the weapon should be substantially higher than that of the new soldier.

These considerations must permeate the weapons review process. First, any state contemplating developing and fielding autonomous weapons must, if it has not already done so, implement a formal weapons review programme. Second, only legal advisors with a high degree of competence in LOAC and international law practice should be detailed to conduct weapons reviews. Third, these legal advisors must be provided the opportunity to develop in-depth understanding of the nature of the artificial intelligence that will be at the proverbial 'heart and mind' of the system. Fourth, the state (normally acting through senior levels of the armed forces) must establish review and validation standards that provide a high level of LOAC compliance confidence prior to fielding the system.

[49] See 'How it happened: massacre in Kandahar', *BBC News* (17 March 2012), www.bbc.com/news/world-asia-17334643.

Reconceiving the concept of command responsibility

This review process, if properly conducted and based on clearly established standards that define the degree of accuracy necessary for validating an autonomous system is capable of LOAC compliance, should produce a sufficient level of confidence in the system to justify fielding and employment. The level of LOAC compliance confidence must, in turn, dictate the extent of permissible lethal force capability fielded in the form of an AWS. In this regard, there is no conceptual justification for imposing a per se requirement that autonomous weapons employ only less than lethal force. Instead, the test for when such weapons should be permitted to employ lethal force should in theory be no different than the test for when a human soldier may employ such force.[50]

Armed conflict involves a contest between organized belligerent groups. The objective of employing force in this contest is to bring about the complete submission of the enemy efficiently through the lawful use of violence. The essential distinction between a use of force in this context and in a peacetime context is that it is the enemy in the collective sense, and not the individual enemy operatives, who is the object of this effort. As a result, use-of-force authority is based on the presumptive threat posed by members of the enemy group and not on individualized conduct-based threat validation. Since such members represent a presumed threat unless and until rendered *hors de combat*, attacking forces are legally justified in employing deadly combat power against such members as a measure of first resort. This protects these forces from the inherent risk of tactical hesitation and serves the legitimate objective of imposing their will on the enemy in the collective sense.[51]

There is no theoretical reason to deprive autonomous weapons of analogous authority. As weapons systems, they, like soldiers, should be

[50] Nor should considerations of 'combatant privilege' be interjected in the analysis of permissible lethal force capability. Indeed, this privilege is simply an irrelevant concept in this assessment, as it relates more to the protection of the lawful combatant from criminal sanction for a lawful conduct during hostilities than it does to the permissibility of engaging in hostilities. An autonomous weapon, unlike a human, is a piece of equipment, and like all other military equipment will be subject to confiscation and either destruction or use by a capturing enemy so long as distinctive markings are properly substituted.

[51] See L. Blank *et al.*, 'Belligerent targeting and the invalidity of a least harmful means rule', available at papers.ssrn.com/sol3/papers.cfm?abstract_ibid=2271152; see also G. S. Corn, 'Mixing apples and hand grenades: the logical limit of applying human rights norms to armed conflict', available at http://papers.ssrn.com/sol3/papers.cfm?abstract_ibid=1511954.

employed in a manner to produce maximum effect on the collective ability of an enemy to resist friendly forces. It is certainly true that, unlike with human soldiers, there is no genuine concern that tactical hesitation may result in mortal danger. However, this is only one aspect of the justification for employing deadly force as a measure of first resort against enemy belligerents. Equally significant is the tactical and strategic effect inherent in such authority. Subjecting enemy forces to the risk of status-based targeting contributes to seizing and retaining operational initiative and setting the tempo of battle. Granting the enemy an automatic reprieve from the full scope of this authority simply because the weapons system employing force is not human provides the enemy with an unjustified windfall and potentially cedes this initiative to the enemy. It also potentially enables the enemy to leverage this constraint to economize a commitment of resources against these assets and mass resources where the enemy believes more robust use-of-force authority may be leveraged.

Of course, as with a human soldier, there may be reasons related to mission accomplishment that warrant restraining the full scope of belligerent targeting authority and imposing an 'attempt-to-capture' restriction on autonomous weapons. As a result, it seems essential that the cognitive capacity of the weapon includes not only an ability to assess when a human subject to lawful attack is rendered *hors de combat* and, hence, protected from further violence but also an ability to assess when such an individual should be offered the opportunity to surrender prior to that point of incapacity. This will enable the employing commander to tailor the weapon's effects to the needs of the mission.

Another issue related to this concern is whether autonomous weapons will ever be able to conduct the nuanced balancing necessary to comply with the targeting principle of proportionality. The impact of this concern may very well turn on the nature of the weapon and will certainly be impacted by the established parameters of permissible use. For example, this would be a minimal concern for a weapon authorized for use only in areas with minimal to no civilian presence, whereas authorizing use of the weapon in a civilian population centre would require a very different cognitive capacity. It should also be noted that the nature of the attack parameters should influence this concern. Thus, an AWS developed as a substitute for human soldiers engaged in close combat will certainly require a capacity to engage in split-second distinction judgments but not necessarily in sophisticated proportionality balancing. The reality of combat indicates that soldiers engaged in close combat cannot engage in

the type of deliberate and sophisticated proportionality assessments that are expected from a battle staff involved in deliberate targeting decisions. Instead, their judgments are by nature swift and ad hoc and will almost inevitably gravitate towards the distinction component of lawful targeting. Perhaps artificial intelligence will actually enable commanders to enhance the proportionality component of this equation, which would be ideal. However, there is no reason to demand more from the autonomous weapon than is demanded of the human soldier.

Developing artificial intelligence capable of this type of sophisticated human judgment may well be impossible. If so, no military or civilian leader may authorize the procurement and fielding of such a weapon. However, once it is determined that an AWS is capable of such efforts, there is no compelling justification for subjecting these systems to a use-of-force regime more restrictive than that imposed on human soldiers. The complete removal of human judgment from the target decision-making process, however, does justify the prioritization of the development phase of fielding over the employment phase as the necessary LOAC compliance focal point. Furthermore, the nature of such potential weapons also necessitates reconsideration of another critical LOAC compliance mechanism: command responsibility. This reconception should, like the process of fielding autonomous weapons writ large, produce a modification or evolution of the focus of command responsibility from the field commander who employs the weapon to the commander responsible for the procurement and fielding decision.

This will undoubtedly present significant conceptual obstacles. The doctrine of command responsibility is obviously premised on responsibility being linked to command and rarely will a procuring official be in a position of command. Even as it is extended to civilian superiors, the doctrine is still premised on the individual possessing directive authority over the subordinates whose LOAC violations are imputed to the leader. Thus, as currently conceived, the foundation of the doctrine will not support the weight of shifting the focus of imputed criminal liability to a procurement official.[52]

However, this need not function as an absolute impediment to such a reconception. Instead, a focus on the underlying rationale of the doctrine suggests that extension to procurement officials, what I call procurement

[52] See Rome Statute of the International Criminal Court (Rome Statute), UN Doc. A/CONF.183/9 (1998); see also Corn, *The Law of Armed Conflict*, 530–1; see generally Hansen, 'What's good for the goose'; Shany and Michaeli, 'The case against Ariel Sharon'.

responsibility, is indeed viable. Ultimately, command responsibility involves two distinct theories of criminal liability. The first is uncontroversial: traditional accomplice liability. Where a commander shares a criminal *mens rea* and acts in a manner (either by commission or omission) that contributes or facilitates an LOAC violation at the hands of a subordinate, liability is attributed to the commander, as it would be to any other individual under the doctrines of accomplice liability.[53]

The second theory subjects commanders to individual criminal liability for the foreseeable LAOC violations committed by subordinates even when there is no evidence that the commander shared the culpable *mens rea* with the subordinate. This is the well-known 'should-have-known' theory of command responsibility: commanders are responsible for LOAC violations that they 'should have known' would occur even if they did not affirmatively encourage or condone those violations. First enunciated in the seminal US Supreme Court *Yamashita* decision,[54] and controversial since that date,[55] this theory of command responsibility creates a powerful incentive for commanders to maintain awareness of the conduct of subordinates and to respond promptly to indicators of misconduct within the ranks. In so doing, the commander discharges his/her responsibility and, in so doing, renders subsequent LOAC violations objectively unforeseeable.[56]

Exploring the complexities of each of these theories of command responsibility is well beyond the scope of this chapter. It suffices that even the more expansive theory is premised on the culpable failure to prevent objectively foreseeable LOAC violations. This is the seed from which an extension of the doctrine to procurement officials should blossom. Indeed, for autonomous weapons, it is the significance of the procurement phase of fielding that makes shifting the liability lens from the employing commander to the procuring official both logical and equitable.

One conceptual obstacle to analogizing autonomous weapons with soldiers is related to the requirement that armed forces operate under responsible command. As noted earlier, key among the responsibilities of

[53] See Rome Statute, Article 28; see also Corn, *The Law of Armed Conflict*, 530–1; see generally Hansen, 'What's good for the goose'; Shany and Michaeli, 'The case against Ariel Sharon'.

[54] See *In re Yamashita*, 16–17. [55] See *ibid.*, 34–41 (Murphy, J, dissenting).

[56] See Rome Statute, Article 28; see also Corn, *The Law of Armed Conflict*, 530–1; see generally Hansen, 'What's good for the goose'; Shany and Michaeli, 'The case against Ariel Sharon'.

command is training and preparing the soldier for the exercise of combat judgments – developing combatants who are capable of exercising autonomous judgment as to what is and what is not a lawful use of combat power. By subjecting commanders to imputed criminal responsibility for subordinate LOAC violations that they 'should have known' would occur, the doctrine of command responsibility creates a powerful disincentive for commanders to fail in this responsibility. This doctrine is therefore integral to the high level of confidence that these human 'weapon systems' will, in fact, properly exercise battlefield judgment.[57]

Accordingly, this doctrine is vital in ensuring that the resources committed to conflict are properly prepared and supervised in order to minimize the risk of LOAC violations. For the soldier, the focal point of this responsibility is the fielding commander. This is logical, as history demonstrates that LOAC violations are usually attributable to unit climate, the nature of orders and directives, and the disciplinary response to indications that the unit is descending into the abyss of disrespect for the law. The role of the commander and the doctrine of criminal responsibility derived from that role provide the genuine lynchpin between fielding armed forces with the explicit task of unleashing deadly combat power and limiting the harmful consequences of that power in accordance with the LOAC.[58]

An analogous concept of criminal responsibility should also become an essential component to ensure AWS comply with LOAC norms. The focus of command accountability must shift, however, from the field commander to the military and/or civilian officials responsible for procuring and fielding these weapons systems. This is a logical outgrowth of the relationship between command responsibility and mitigating the humanitarian risks associated with autonomous weapons. Modifying the doctrine to apply to procurement officials, and not only to fielding commanders, will emphasize that it will ultimately be decision-making officials, and not technicians or legal advisers, who must validate the capability of the emerging technology. Knowing that they bear responsibility for objectively foreseeable technological failures should make these decision makers cautious to field such capability, which will, in turn,

[57] See Rome Statute, Article 28; see also Corn, *The Law of Armed Conflict*, 530–1; see generally Hansen, 'What's good for the goose'; Shany and Michaeli, 'The case against Ariel Sharon'.

[58] See Rome Statute, Article 28; see also ICRC, *AP I Commentary*, 1017–23; Corn, *The Law of Armed Conflict*, 530–1; see generally Hansen, 'What's good for the goose'; Shany and Michaeli, 'The case against Ariel Sharon'.

mitigate the risks associated with taking human judgment out of the actual target engagement decision.

Accordingly, these officials must understand they will be accountable for objectively foreseeable failures of the weapon review and compliance validation process: if a fielded weapon produces a LOAC violation, and it is determined that the procurement process was objectively insufficient to ensure LOAC compliance, the official who approved the weapon will bear responsibility for the violation. This process should logically result in demands for the highest level of confidence that the nature of the weapon system is capable of effectively implementing obligations of distinction, proportionality and precautionary measures if and when employed.

In short, not just the employment but also the development and procurement of autonomous weapons must fit squarely within this complex interrelationship between the lawful combatant status – the status that provides the international legal privilege to participate in hostilities – and the requirement for such participants to operate under responsible military command. Lawful combatant status is established by compliance with the four criteria incorporated into the Third Geneva Convention's prisoner of war qualification provision, which collectively indicate the link between the responsibility of commanders and the legally sanctioned exercise of combatant judgment.[59] Indeed, it is the influence of responsible command on subordinates that distinguishes 'combatant' – individuals authorized to engage in hostilities – from others who, although entitled to prisoner of war status upon capture, are not vested with this authority.[60]

The importance of the link between responsible military command and combatant qualification is not a novel concept. Requiring subordination to responsible command as a condition for qualifying as a

[59] Geneva Convention Relative to the Treatment of Prisoners of War (Third Geneva Convention) 1949, 75 UNTS 135.

[60] This dichotomy is revealed by considering the extension of prisoner of war status to civilians who provide field support to the armed forces. Three of the four combatant qualification requirements could easily and routinely be satisfied by these civilians: carrying arms openly; wearing a fixed and distinctive symbol and complying with the LOAC. But these individuals, although entitled to prisoner of war status upon capture, are not considered combatants within the meaning of the law. The true distinguishing factor between members of the regular armed forces and associated militia groups, on the one hand, and civilian augmentees, on the other, is therefore operating within the context of the type of command relationship that is essential to ensure LOAC compliance: operating under 'responsible command'.

combatant is tethered back to the 1899 and 1907 Hague Regulations.[61] This connection was finally codified in 1977 when Additional Protocol I defined the term combatant. Article 43 of Additional Protocol I explicitly establishes the existence of military command and discipline as a condition for recognizing combatant status:

> Article 43 – Armed forces
>
> 1. The armed forces of a Party to a conflict consist of all organized armed forces, groups and units which are under a command responsible to that Party for the conduct of its subordinates, even if that Party is represented by a government or an authority not recognized by an adverse Party. *Such armed forces shall be subject to an internal disciplinary system which, 'inter alia', shall enforce compliance with the rules of international law applicable in armed conflict.*[62]

The rationale for this requirement seems clear: in order to ensure compliance with the LOAC, only those individuals subject to military command and discipline should be permitted to perform functions involving the type of discretion that, if abused, might result in LOAC violations. This is reinforced by the commentary to Article 43:

> This requirement [the link between combatant status and internal command discipline and control structure] is rendered here with the expression 'internal disciplinary system', which covers the field of military disciplinary law as well as that of military penal law . . . The principle of the inclusion of this rule in the Protocol was from the beginning unanimously approved, as it is clearly impossible to comply with the requirements of the Protocol without discipline . . . *Anyone who participates directly in hostilities without being subordinate to an organized movement under a Party to the conflict, and enforcing compliance with these rules, is a civilian who can be punished for the sole fact that he has taken up arms, unless he falls under one of the categories listed under (2) and (6) of Article 4A of the Third Convention (categories (1) and (3), which cover the regular armed forces, should automatically fulfil these requirements).*[63]

The commentary omits any reference to individuals entitled to prisoner of war status by operation of Article 4A(4) of the Third Geneva Convention – civilians who support the forces in the field. The significance of this omission is obvious: the definition of combatant – individual entitled to participate in hostilities – conclusively presumed that

[61] See L. Green, *The Contemporary Law of Armed Conflict* (Melland Schill Studies in International Law), 2nd edn (Manchester University Press, 2000), 102–9.
[62] See Additional Protocol I, Article 43 (emphasis added).
[63] See ICRC, *AP I Commentary*, 513–14 (emphasis added).

individuals not fully incorporated into the military command-and-control structure (not 'part of the armed forces') could not engage in combatant activities. As Michael Schmitt notes:

> There are but two categories of individuals in an armed conflict, combatants and civilians. Combatants include members of a belligerent's armed forces and others who are directly participating in a conflict. As noted, the latter are labeled unlawful combatants or unprivileged belligerents; they are either civilians who have joined the conflict or members of a purported military organization who do not meet the requirements for lawful combatant status. Everyone else is a civilian, and as such enjoys immunity from attack.[64]

In Schmitt's continuum, civilian support personnel may not be considered combatants. This conclusion reflects the LOAC's essential linkage between being incorporated into a military unit, subject to responsible command, and the legal privilege of operating as a combatant. This link serves the interests of LOAC compliance by emphasizing to the military commander – the individual with the most direct and meaningful opportunity to ensure respect for the law – that violations jeopardize not only state and international interests but also the commander's personal interest of avoiding imputed criminal responsibility for foreseeable subordinate misconduct. It also suggests that legitimate combatant status is contingent not simply on whether or not an individual will take a direct part in hostilities, or wears distinguishing clothing and equipment, but, instead, on the expectation that the individual is subject to the fundamental compliance mechanism of the LOAC – a military command–subordinate relationship. Thus, for the human warrior, the true *sine qua non* for determining the limits on authority that may be exercised in armed conflict is whether performance of the function requires an exercise of judgment implicating LOAC compliance.

This connection also reveals that the LOAC has historically relied on the loyalty and discipline inherent in the command–subordinate relationship, bolstered by the proscriptive and disciplinary authority over combatants, to maximize confidence that individual human actors exercise judgment in accordance with the law. It is therefore no surprise that operating under responsible command is an essential element for qualifying as a lawful combatant. Military command authority over subordinates is relied upon to emphasize compliance obligations and ensure proper LOAC training

[64] See M. Schmitt, 'Humanitarian law and direct participation in hostilities by private contractors or civilian employees', *Chicago Journal of International Law*, 5 (2004–5), 522.

and individual preparation for the exercise of combatant judgment.[65] Being a member of a military unit is also expected to produce a high degree of loyalty to the military commander, often referred to as the concept of unit cohesion. This bond of obedience and loyalty is a unique and critical aspect of military organizations. It is unquestionable that the military command relationship, to include the proscriptive and disciplinary authority inherent in that relationship, is an essential element in the scheme of LOAC compliance mechanisms.

Subjecting commanders to criminal responsibility for LOAC violations that they 'should have known' would occur enhances confidence that individual soldiers will comply with the LOAC.[66] Subjecting commanders to imputed criminal liability for foreseeable subordinate LOAC violations that did not evolve as an aberration ensures, instead, that commanders diligently execute their responsibilities to ensure subordinate LAOC compliance. These responsibilities include training subordinates in their legal obligations; involving legal advisers in operational decision making; establishing a command atmosphere that emphasizes good faith compliance with the law and taking swift disciplinary action in response to any breach of the law.[67]

This relationship between responsible command and the exercise of human judgment is logical and critical as a LOAC violation risk mitigation measure. However, when pre-established artificial intelligence dictates the exercise of judgment, this relationship cannot have its desired effect. It is therefore essential to ensure that the application of the doctrine of command responsibility focuses on the command level with genuine 'responsibility' for the exercise of judgment. This level is not the field or employing commander but, rather, the procuring commander. Only by emphasizing that it is this level of command that is subject to imputed criminal responsibility for objectively foreseeable failures of autonomous weapon LOAC compliance capability will the rational of the doctrine properly align with this emerging technology.

The intangible 'force multiplication' effect of the human warrior

The conclusion that truly autonomous weapons will evolve to a point where they may be employed with a sufficient degree of LOAC

[65] See Green, *The Contemporary Law of Armed Conflict*, 280–6.
[66] See generally *In re Yamashita*.
[67] See Green, *The Contemporary Law of Armed Conflict*, 277–83.

compliance confidence undermines a purely humanitarian-based demand for a per se prohibition. However, this conclusion does not indicate that military leaders will perceive employment of such weapons as desirable. Ultimately, such employment will depend on both the validation that these weapons comply with LOAC obligations and the determination that the contribution they make to effective military operations justifies this next step in the evolution of the means of warfare.

It is not self-evident that this will be the case. Conceptually, fielding the truly autonomous weapon would result in a quantum leap in military affairs. While military organizations seem eager to embrace technology as a supplement to human action, the notion of using technology as a substitute presents very different considerations. This is because such a development would be truly inconsistent with the perception of the well-trained soldier, capable of engaging in human reasoning to navigate the most complex battlefield decisions, as the most valuable and effective weapons system available for a commander to employ.

Military commanders understand perhaps better than anyone that well-trained soldiers led by quality leaders will often prove decisive in combat. This is not because these soldiers are viewed as 'robots' who simply follow orders without question. Instead, it is the sophisticated reasoning of soldiers that equips them to lead and follow effectively in a manner that advances mission accomplishment. In fact, the very concept of 'mission command' – a concept that is absolutely central to the planning and execution of US (and other nation's) military operations – is premised on the expectation of subordinate initiative to advance the commander's intent. US Army Doctrinal Publication 6–0 (GDP 6.0) is devoted to this concept and defines mission command as 'the exercise of authority and direction by the commander using mission orders to enable disciplined initiative within the commander's intent to empower agile and adaptive leaders in the conduct of unified land operations'.[68] More specifically, mission command is explained as follows:

> The exercise of mission command is based on mutual trust, shared understanding, and purpose. Commanders understand that some decisions must be made quickly at the point of action. Therefore, they concentrate on the objectives of an operation, not how to achieve it. Commanders provide subordinates with their intent, the purpose of the operation, the key tasks, the desired end state, and resources. Subordinates then exercise

[68] US Department of Army, *Mission Command*, Doctrine Publication no. 6–0, 17 May 2012; C1, 10 September 2012; C2, 12 March 2014, para. 2, available at http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/adp6_0.pdf.

> disciplined initiative to respond to unanticipated problems. Every Soldier must be prepared to assume responsibility, maintain unity of effort, take prudent action, and act resourcefully within the commander's intent.[69]

The publication then explains that:

> An effective approach to mission command must be comprehensive, without being rigid, because military operations as a whole defy orderly, efficient, and precise control. Military operations are complex, human endeavors characterized by the continuous, mutual give and take, moves, and countermoves among all participants. The enemy is not an inanimate object to be acted upon. It has its own objectives. While friendly forces try to impose their will on the enemy, the enemy resists and seeks to impose its will on friendly forces. In addition, operations occur among civilian groups whose actions influence and are influenced by military operations. The results of these interactions are often unpredictable – and perhaps uncontrollable.[70]

Then, in an explicit invocation of the significance of human judgment in the implementation of mission command, the publication provides:

> A HUMAN SOLUTION TO COMPLEX OPERATIONAL CHALLENGES
>
> To overcome these challenges, mission command doctrine incorporates three ideas: the exercise of mission command, the mission command philosophy, and the mission command warfighting function. In this discussion, the 'exercise of mission command' refers to an overarching idea that unifies the mission command philosophy of command and the mission command warfighting function – a flexible grouping of tasks and systems. The exercise of mission command encompasses how Army commanders apply the foundational mission command philosophy together with the mission command warfighting function. The principles of mission command guide commanders and staffs in the exercise of mission command.[71]

It is difficult to imagine a more emphatic indication of the importance of human judgment in the execution of military operations, not merely because such judgment is important for the implementation of LOAC obligations but also because it is central to the effective execution of a commander's intent. It is very likely that effective commanders will be reluctant to embrace technological substitutes for human subordinates

[69] *Ibid.*, para. 5. See generally General M. E. Dempsey, 'Mission command', *Army Magazine*, 61 (43) (2011), 43–4 (discussing both philosophical and definition changes for what defines mission control), available at www.ausa.org/publications/armymagazine/archive/2011/1/Documents/Dempsey_0111.pdf.

[70] Doctrine Publication no. 6–0, para. 3. [71] *Ibid.*, para. 4.

not merely because of concerns related to LOAC compliance but also perhaps because of the more significant concern that these substitutes cannot function with the type of human intellectual nuance necessary to fit within this mission command model.

Indeed, the rationale for asserting a shift from command responsibility to procurement responsibility – that fielding commanders will be incapable of asserting any meaningful influence on the cognitive reasoning of these weapons – may be the operational Achilles heel that results in the hesitation to pursue them. This is because developing a relationship of trust and confidence between commander and subordinate is central to mission command – or perhaps even the broader concept of 'responsible command'. As Doctrine Publication no. 6–0 notes, building 'cohesive teams through mutual trust' is the first principle of effective mission command and:

> [e]ffective commanders understand that their leadership guides the development of teams and helps to establish mutual trust and shared understanding throughout the force. Commanders allocate resources and provide a clear intent that guides subordinates' actions while promoting freedom of action and initiative. Subordinates, by understanding the commander's intent and the overall common objective, are then able to adapt to rapidly changing situations and exploit fleeting opportunities. When given sufficient latitude, they can accomplish assigned tasks in a manner that fits the situation. Subordinates understand that they have an obligation to act and synchronize their actions with the rest of the force. Likewise, commanders influence the situation and provide direction, guidance, and resources while synchronizing operations. They encourage subordinates to take bold action, and they accept prudent risks to create opportunity and to seize the initiative.[72]

This brief discussion of the relationship between human judgment and mission command is really just the tip of the proverbial iceberg in relation to the significance of human reasoning as an essential element of effective military operations. However, even this cursory discussion illustrates a critical point: analogizing an autonomous weapon to a human soldier for the purposes of assessing LOAC compliance confidence is not intended to indicate that technology and humans are simply interchangeable. There are other aspects of effective military operations where even the type of conceptually advanced artificial intelligence that could produce LOAC compliance confidence would be insufficient to justify treating the autonomous weapon as a human substitute.

[72] *Ibid.*, para. 6.

Opponents of these weapons may therefore be pleasantly surprised by a lack of enthusiasm for these weapons among military leaders; perhaps not for the reasons they advocate, but because of an unwillingness to entrust mission execution to anything other than a human subordinate.

Conclusion

Whether it is possible to develop the sophisticated artificial intelligence needed to enable a weapon to operate with a level of cognitive reasoning analogous to that of a human soldier is yet to be seen. However, it is almost inevitable that attempts to achieve this goal will gain momentum. If this comes to fruition, there is no normative infirmity with allowing the use of such weapons during armed conflict. Accordingly, instead of pursuing a per se prohibition against autonomous weapons, core concepts related to ensuring human actors comply with LOAC obligations must be adjusted to address the realities of these weapons. Doing so will preserve the historic animating purpose of the law: facilitate military mission accomplishment while mitigating the human suffering associated with war.