Leadership challenges arising from the deployment of lethal Autonomous Weapon Systems (AWS)

How erosion of human supervision over lethal engagement will impact how commanders exercise leadership

Dr Paddy Walker
April 2020

Introduction

Autonomous Weapon Systems (AWS) are defined as robotic weapons that have the ability to sense and act unilaterally depending on how they are programmed. Such human-out-of-the-loop weapon platforms will be capable of selecting targets and delivering lethality without any human interaction. This technology may still be in its infancy but both semi-autonomous and other pre-cursor systems are already in service. There are, moreover, several drivers to a move from merely automatic weapons to fully autonomous weapons which are able to engage a target based solely upon algorithm-based decision-making. This requires material step-change in both hardware and software and, once deployed, posits a significant change in how humans wage war. Notwithstanding that complex technical difficulties must first be overcome, the introduction of AWS also pose basic legal, moral and ethical challenges.

The purpose of this paper is to consider the wider challenges on leadership from the deployment of such weapons and how unsupervised engagements might degrade the commanders' craft.[1] The piece concerns unsupervised, autonomous lethal engagement. Beyond its scope are the many appropriate uses of battlefield autonomy such as in logistics, intelligence gathering, security and other repetitive deep, dull, dirty and dangerous tasks. Nor is it about operational benefits gained by removing human intervention such as force multiplication, ethical governance and tactical considerations. In considering the decision to cede control of killing to an algorithm, the paper intentionally amalgamates the several levels of command and other relevant parties that will together comprise AWS leadership into a single entity which it terms the 'Delivery Cohort'. This is therefore a broad grouping intended to reflect what will be a distributed deployment responsibility. The Cohort encompasses politicians and generals, local commanders as well as soldiers on the ground. It includes maintenance and service personnel, those in manufacturing and procurement, in design and programming, those heading commercial entities, the regulators and lawyers, the Press and those from the Third Sector. As a broad proxy for the deployment process, the construct also merges together the disparate constituencies that then influence that Cohort, from complex rules of engagement to the trials of deploying AWS in multi-service, multi-force and multi-national roles. The paper generally considers the leadership element from the perspective of parties that are immediate and local to the use of AWS.[2] In looking to identify key themes, the paper also concentrates upon the deployment of broad-use, wide-capability AWS.

In considering leadership in an age of unsupervised lethality, certain generalities arise. First, there are several plausible drivers pushing weapon design towards weapon autonomy. These include runaway budgets, both in a State's capital account but also in its day-to-day outlay running modern manned battle. They also include issues of political expediency, the prospect of removing soldiers for front-line harm while delivering surgically precise lethality, advantages from increasing operational speed, of remote delivery and force multiplication. These characteristics underpin what commentators term a broad *revolution in expectation* in AWS deployment.[3] It is therefore unsurprising that leaders will be drawn towards lethal autonomous engagement. This paper aims to ensure that leaders are aware of this argument's verso: Unsupervised weapons cannot themselves make legal

determination and it therefore remains the responsibility of AWS' Delivery Cohort to ensure legal use of these assets.

What constitutes military leadership in AWS deployment?

It is useful from the outset to map the elements of leadership under discussion. One workable proxy is provided by the *Eleven Principles of Leadership* first developed in 1948 and subsequently published in the US Army's Field Manual on Leadership. Sixty years later, its principles are still used across the UK Armed Forces.[4] Its code is surprisingly monotonal. *All* of the principles are behavioural, human and relate almost exclusively to context. As a corpus, it seeks to define the changeable art of motivating a group of individuals to act toward achieving a common goal. The leadership paradigm that it outlines is therefore an abstract. It is volatile and defies prescriptive description. This is a key observation as it suggests strongly that leadership cannot then be captured by coding strings in the AWS' algorithm.

The *Principles*' repeated notion that leaders should 'seek responsibility and take responsibility for their own actions' reinforces this irreducibility in which the deployment of independent, out-of-the-loop munitions cannot readily be accommodated. Indeed, it is difficult to fit *any* of its eleven guidelines into a leadership narrative based upon AWS deployment. Leaders, for instance, must lead by example if they are to expect courage, responsibility, initiative, competence, commitment and integrity in their subordinates. *Principle 6*, 'keeping your people informed', underscores the empiric that a chain of command performs best when it well understands the *context* in which it is operating. Here, commanders should therefore be wary of circumstances that must dislocate that chain's information flows. Given that successful tasks are understood, supervised and monitored, deploying assets that lie outside the Cohort's supervision must weaken the cycle of intelligence, decision-making and reporting. Finally, the *Principles* note the importance of making *timely* decisions; good decisions made at the right time are better than the best decisions made at an inappropriate time. In this manner, indecision (here, perhaps the hesitancy experienced by a leader on exactly how any independent weapon will act) creates general hesitancy, loss of confidence and confusion.

If the tenets of leadership cannot be reduced into code-ready packets, the Cohort must still decide what it 'means' today to be a leader and how incorporating weapon autonomy changes this model. How does the commander reconcile ceding the decision to kill to an autonomous agent (that is independent of that command) with the leadership traits of, say, on-the-ground bravery, initiative and example? The second issue relates to the nature and extent of how independent weapons are embedded in that battlecraft. What role might be appropriate, how fit-for-purpose is the AWS technology for deployment and what degree of dilution in legal compliance is to be tolerated in the use of AWS by the deploying State. A commander's decision-making, after all, is knowing *if* to decide and then *when* and *what* to decide.[5] The process currently hinges on clear articulation of the commander's battlefield visualisation. The complexity of that task is best evidenced by the process' currently accepted tools including standard operating procedures, exhaustive field manuals and granular rules of engagement. In considering the degree of change that AWS will occasion, current practices at least provide a benchmark against which to compare what

might be unsupervised processes. Combat decisions today are generally the product of rigorously formal courses of action (COA) in order to distil best decisions through a process of assessments and approvals. They require articulation of the mission, agreement on planning timelines, intelligence preparation, staff estimates, rehearsals and analysis of asset cohesiveness, experience and conditioning. In order to posit solutions, current planning involves definition of tactical issues and their relevant constraints, the articulation of critical assumptions and risk assessment as well as wide consideration of adjacent factors. Current planning requires review of contingencies, of fire support guidance, of mobility and counter-mobility support and myriad other security measures. It is the gap between manned intelligence and those same requirements under unsupervised engagement that leads Scharre to note a 'valley of death between R&D and operational use'.[6]

Before initiating AWS deployment, the commander must also factor for the several existing limits under the rules of International Humanitarian Law (IHL). It is long-established that rules on distinction, proportionality and precautions-in-attack must be complied with by all persons who plan, decide upon and carry out attacks. Such assessments under IHL, however, involve evaluative and contextual judgements for which the local commander remains both responsible and accountable. Under existing IHL, commanders deploying AWS must retain sufficient human control to enable context-specific judgement as required under such rules. commanders, for instance, must be confident that their unsupervised assets direct attacks only at military objectives which contribute to military action and whose destruction, capture or neutralisation offers definite military advantage. In this vein, the commander must also be satisfied that incidental civilian damage should be proportional to the advantage anticipated. Furthermore, IHL requires that the commander has *specific* knowledge of the context and circumstances prevailing at the time of the attack prior to authorising that attack. This includes *each* targets' specifics and circumstances as defined within the context of such weapons' overall effects. In practical terms, the commander must therefore require a minimum level of system predictability across their entire environment of use. Only then can the commander be confident on the degree of constraint that is appropriate for AWS' tasking, its targets and deployment environments as well as the spatial and geographical boundaries of every ensuing lethal engagement. Compliance with this list is already complicated in manned warfare. Deployment of self-learning weapons and the requirement that a commander's planning assumptions (and legal assessments) remain valid *throughout* each attack undermines the compliant deployment of AWS.

AWS' behavioural issues and the role of the combatant

In the first instance, fielding independent weapons certainly challenges traditional notions of what it is to be a combatant. As noted by Enemark, the longbow was outrageous at the time because of the new degree of force that it brought to the battlefield but also because the lowly peasant could now defeat the aristocrat.[7] In this vein, the drone operator already kills without any material personal risk. Autonomous weapons promise an even higher degree of user insulation. In instances of unsupervised engagement, where does accountability really rest in what has now become an *independent* attack? Enemark terms the notion ‘post heroic warfare’, questioning the traditional construct of Nation States prepared to wage war on the basis of substantial casualties.[8] Any shift to independent

engagement might prompt an even wider gulf between civilian and military values, a disappearing paradigm of noble death and the rise of a new norm which is based instead on remote warfare, few combatants and fewer casualties.

This certainly impacts on notions of leadership. For those in battle, several attributes have long characterised combat soldiers across cultures and time. Endurance, courage, strength, skill and honour. Riskless war would dilute these requisites of valour. No longer need the leader be admired, respected, cheered (even mourned). Removing human oversight must therefore blunt traditional notions of duty and self-sacrifice, the covenants binding soldiers together and to the society that they serve as well as the ageless notion of subjugating self-preservation in the present to make life better in the future. Riskless warfare, first the unmanned drone but soon the unsupervised weapon, certainly ends the concept of a soldier having a contract to kill that is based upon a risk of being killed. Context is provided here by understanding emerging language and processes; a drone, for instance, in colloquial Urdu is a derogatory term suggesting cowardice for sending robots to do man's work. In the same vein, US Air-Force language of 'bugsplat' seemingly belittles collateral damage in a remote strike.

Does the AWS deployment model alter the analysis?

There is also clearly a sliding scale to the adoption of AWS. Gradual replacement of supervised componentry by autonomous agents as one or more parts of an overall weapon system now requires the local commander to understand boundaries to his control and where such jurisdiction has been ceded to a machine agent. Further along this continuum might then see machine-human teaming where the experience of human soldiers is leveraged by companion autonomous technologies. Such hybrid autonomy involves the toggling of command between human and machine and presents a new set of leadership issues. Indeed, is flexible autonomy practically feasible? Research, after all, uniformally points to erratic performance when humans are required to intervene in moments of high stress or in situations of limited information. Moreover, the human has hitherto been operator, moral agent and failsafe across an engagement's whole decision waterfall. It has been the human operator who has managed engagement intangibles such as task complexity, its cognitive workload, its number, type and duration. The command issue here is how this equation is to be managed in the dynamic state of hybrid operations.

Hybrid autonomy will also complicate the commander's understanding of tasking which must be completed and then communicated in *advance* of each mission. Successful outcomes are empirically based upon commanders being able to recognise when metrics change within that mission. Tasking in today's environment is already complicated by matters of scale, margin of error, task space and colleague assets. As weapon autonomy is introduced, commanders must factor for asset independence and style drift while slack time for battle processes is reduced, scope for rule-bending and initiative is removed and, in the case of weapons based upon machine learning (ML), the leader's ability to predict and influence outcomes is lessened.

Technical challenges in deploying AWS?

An autonomous weapon must work first-time, every-time. It must do this to be reliable, to fulfil the expectations of its broad Delivery Cohort and to adhere to legal frameworks. Its operation must be predictable and meet moral and ethical standards. To justify its deployment, the Cohort must be confident that the AWS will perform its defined task and undertake this better (whether that be cheaper, quicker, more consistently, more surgically) than available manned alternatives. Commanders, after all, have deployment and allocation choices. It is for this reason that AWS efficacy boils down to the weapon's technical basis and its fitness for purpose.

How then is it envisaged that human supervision can appropriately be removed from weapon operation? Notwithstanding the pace of technical innovation, precepts around the AWS' ML framework must certainly be understood by the commander. ML's intended role is to enable the weapon to undertake on-board deduction and independent prediction whereby whatever the weapon has experienced in prior cases should inform what the weapon should expect now. In the context of AWS deployment, it is to process what is a known case whose relationships can then be carried over to the present case. This, however, requires the consolidation of a very broad portfolio of technologies, each requiring bespoke configuration in order to mediate contradictions and evaluate significance while rejecting alternatives that might occasion unsatisfactory outcomes in weapon decision-making. Further technical disruption is required across machine processes before this can be considered realistic.[9]

The overarching architectural intention of AWS design is that the weapon's artificial neural network learns by example and this therefore requires that commanders factor for the several characteristics that comprise its core applications. These include pattern recognition, label matching, data classification and all of the associated routines that will make up the conversion of sensed inputs into weapon actions. This is, fundamentally, a fragile set of arrangements. AWS function will require that each neuron has its own summation and threshold functions. Should a battlefield signal then exceed a threshold (as defined by the Cohort's configuration phase) then it is propagated forward on to other neurons in order to instigate a known action. The weapon might therefore appear 'independent' but is reliant upon human-set configuration in order that it be trained rather than simply programmed. Tolstoy notes in *War and Peace* that the real general 'never finds himself at the beginning of some event'. Instead, the commander is perpetually situated in the middle of a *series* of events, each a link in what is an endless chain of causation.[10] The analogy is useful as it reinforces this enduring infeasibility of trying to hard code rules of engagement. Matching weapon instructions to the complexities of battle require programmers to engineer unrealistic start and end points in order to code appropriate commands. Each coding choice will be constrained by earlier choices. Unless AWS capabilities are very restricted, critical pathways will go undiscovered, ignored or misunderstood in a manner suggesting that decisions tree models make an inappropriate framework upon which to remove supervision.[11] Absent human oversight, the weapon's decision must instead be an overly simple value-maximising exercise that is unacceptably fragile in situations of uncertainty, when priorities change or when underlying instructions require resetting or reconfiguration.

In order to consider leadership ramifications arising from this architecture, the framework requires context. The very largest current artificial neural network today has some sixteen million neurons. This is a laboratory example and equates perhaps to the cognitive function of a small frog. By way of reference, the human brain utilizes some hundred billion such neurons, evidence that true AWS deployment currently remains a remote possibility. A key challenge to AWS' introduction is the application and management of prioritising weights that will be required to encourage intended weapon behaviour and without which its neural network is unable to derive meaning from newly sensed inputs drawn from the weapon's immediate environment. Propagation is therefore a core function whereby the weapon's neurons must be dynamically coached in order to encourage specific behaviours towards specific battlefield outcomes. For this to work, however, the model requires that each data string be labelled and then processed on a near identical basis to that data which the weapon has previously encountered or has stored as an initial representation. It is this characteristic that underpins the acronym CACE whereby *Change Anything and you Change Everything*.[12] Research demonstrates that the introduction of new parameters or slightly heterogenous data to that under which the weapon has been trained will empirically confound AWS' ML processes for the tasks envisaged, particularly in the dynamically changing nature of a battlespace with its prevalence of hidden, partially observable or camouflaged states.[13]

Other challenges arise from the systemic position of ML within AWS applications. The number of datapoints comprising an engagement sequence is, for instance, intractably large (the section in the *Joint Service Manual of the Law of Armed Conflict* that deals just with the conduct of hostilities runs for more than fifty pages[14]) and represents a material step-up in machine processes from anything currently undertaken in the research laboratory.[15] Here, the management of the AWS' neurons will be governed by an error function with the goal of training being to minimise this function. Such a model, however, confines weapon learning to be an approximator and, as above, commanders need to be aware that this is likely inappropriate for compliance with the Laws of Armed Combat (LOAC). The envisaged processes also rely upon seamless data. Here again, research suggests that model performance may even plateau according to the frequency that an autonomous agent polls its sensors, making an ever-smaller change to the model's weights with every new iteration.[16] Adding new training parameters in order to broaden the AWS' data capture empirically results in extra layers of non-linearity which will further challenge optimisation of the weapon's learning routines. A marginally different setup or a marginally different training dataset will lead to very substantial variation in that weapon's decision-making. ML may similarly tend to ignore sensed features that comprise only a small number of examples in a training set but, in the passage of an engagement, ones that possibly account for features of critical battlefield importance.

In this vein, it is also intrinsically challenging for the AWS' sensors, the sole source of inputs into its subsequent decision processes, to garner *consistent* information.[17] Smoke, reflectance, image echo as well as issues of data intensity and saturation all comprise practical difficulties that may enduringly compromise AWS' data-labelling and data-matching models. Similarly, class boundaries separating different data examples resist definition where that data is noisy and indistinct, making it impossible to designate data strings for further statistical analysis. Data mismatch against its training set, anything that is

statistically out of the ordinary, will also confound its unsupervised processes, whether the result of simple feint, by enemy surprise or by this inadequate data separation. Leaders should therefore be concerned by the model's sensitivity (termed its detection rate) and its specificity (here, its false alarm rate). External mediation is by definition difficult and the Cohort must therefore defend against the condition whereby incremental supervision does not then introduce either bias or over-fitting to weapon data.

Special and cognitive considerations of AWS deployment?

A further leadership quandary is that the unsupervised weapon cannot be sure it has found its best action for each state until it has tried all possible actions in all possible states. The combinatory requirement that this posits will be infeasibly large. Current ML models are empirically too iterative to manage the parameter-rich procedure that is target selection, the more so without any appropriately definable end-state. Error, moreover, in the weapon's sensing of its current state will presumably carry forward in that machine's future learning and future battlefield actions. Similarly, much of what the weapon has recently learnt may be invalid if its environment or its combat task changes, highlighting the issue of timing what might otherwise become a perpetual learning phase for the AWS, the trade-off between a weapon that is 'constantly learning' as opposed to one that is using what is already known to work at the cost of missing out on further improvement. Indeed, the challenge of incorporating 'un-learning' into machine routines remains untested and points to a key inefficiency in the preparation of that weapon's input data: As much as two thirds of data analysis is currently taken up with data preparation, an inappropriate luxury on a quick-changing battlefield if the weapon is to be properly valuable yet still compliant.

What then should be expected of an unsupervised weapon's cognitive load?[18] After all, the theory is that appropriately autonomous function should comprise very many processes including knowledge and memory management, attention and evaluation as well as problem solving and decision making. Without just one of these elements, is the weapon deployable and sufficiently robust to replace the human soldier? The difficulty, of course, is that cognition, even in its human condition, is challenging to define. It can be conscious or unconscious, concrete or abstract. It can also be intuitive (here, the 'knowledge' of a task) or conceptual (the 'model' of a task). As above, the construct for AWS deployment is that its cognitive processes must use existing knowledge in order to generate new knowledge. Several leadership issues arise from this requirement. Given that AWS' ML model is based upon data collection and preparation then labelling and matching, current models lack sufficient scripting differentiation to enable the system performance that is envisaged by the Cohort. Weapon datasets must also be incrementally built as not all data is available to the weapon at the same time. Information arrives in waves and any wide-capability AWS model will therefore require seamless data prioritisation and allocation in order to create relevant datasets. This process also requires an appropriate means to manage data denial. Little transferable research is going on around these central tenets.[19] Finally to this point, the weapon's data distribution must evolve over time from the weapon's first deployment and this contradicts AI's fundamental hypothesis of 'identically distributed data' upon which ML and classic data-mining algorithms rely.

Leadership challenges arise from the translating of current manned practices into protocols to govern autonomous weapons. At the very least, leaders must be aware that the simplification methods which characterise human decision-making are an inappropriate start point for the programming of independent weapons. An analysis of such traits is therefore useful. As currently envisaged, AWS operation will be based upon system training using large sets of training data. Here, methods reflect human experience: When faced with new circumstances, combatants will naturally compare them to similar situations previously encountered and now lodged in their memory. For the soldier, however, resulting decisions are rarely the product of thoughtful deliberation, especially in a time-constrained environment. Humans rely upon shortcutting heuristics that confound any simple translation of manned to unmanned function. Most heuristic challenges in AWS programming arise from dealing with the weapon's extensive search set. As uncertainty arises, whether from enemy action or feint, the independent weapon must piece together relevant intelligence without outside agency unless that system is managed to a very narrow remit. It is, after all, enduringly difficult to code a human's imaginability bias whereby soldiers backfill intelligence gaps into a subjective premonition that compensate for when memory and data on previously relevant instances are absent.

The Cohort must also consider other heuristics. In manned engagements, for instance, decision probabilities may be influenced by the degree of danger supposed for an action sequence. In autonomous engagements, decision processes may instead be compromised by *illusory* correlation that then paralyses the weapon as it iterates in attempts to correct for incomplete information. Leaders should also factor for confirmation bias whereby the weapon erroneously validates data links for further processing or, more likely, apportions inappropriate weight to the fit of sensed information and its original dataset. Research demonstrates here that it will be problematic to have the weapon reliably classify that an event or object falls into a particular category of events or objects; in order to categorise a new battlefield occurrence in a timely fashion, the system will instead review and dissect it for characteristics that represent much larger groupings of pre-existing occurrences.[20] Once that datapoint has been tagged to represent the traits of such broader categories, the weapon will action sequences according to what will be an inappropriately narrow class of occurrences. Such representativeness prompts other potential biases including insensitivity to prior probability of outcomes, base-rate neglect, insensitivity to sample size, misconceptions of chance and, generally, failure to identify regression to the mean.[21] Leaders must recognise that the AWS' ML is likely to judge an event more likely as it uncovers further information that its deployment parameters compute relevant to its decision processes: revealing additional detail to a battlefield situation may, in programming terms, make that scenario appear more plausible yet the Cohort must understand that mere discovery of additional information does not affect the probability of any particular situation actually occurring.

Coding constraints to AWS deployment

The fundamental leadership issue here is therefore whether it can ever be appropriate to cede the decision to kill to an algorithm and whether the local commander can reasonably allocate tasks within a battleplan to weapons that cannot be overseen. Is disruptive hardware more relevant that the training and ingenuity of the human soldier given that

sophisticated electronics might actually be confounded by quite low tech solutions. This possibility underpins the theory of nullification that forecasts short-lived advantage from new military technology.[22] The decision, moreover, must take place at a point where future weapons will undoubtedly comprise an increasing number of self-governing sub-parts, each of which are capable of independent agency. Commanders must therefore be aware that each and every weapon component may be capable of autonomy in its own right. Such composite agency is complicated precisely because action selection now depends not only on the impulses of its constituent sub-agents but also on how those sub-agents are organised and coded.

Indeed, AWS' fitness for purpose is compromised by systemic coding challenges.[23] It is coding, after all, that must express all the intentions and constraints of the Delivery Cohort. Only through coding can the weapon's goals and action selection be effected. The leadership challenge here is that the code-based foundation of AWS operation is reliant upon frequent (possibly continuous) scaling factors and conviction weightings to manage the weapon's sensed data. Moreover, the coding model envisaged means that sensed data with least variance will be given additional weighting in each new polling iteration. This has consequences. Confidence weightings have several unintended effects including data smoothing whereby data (in the case of an engagement sequence, information on target signature, classification, location, threat and sensitivity to battlefield clutter) becomes inappropriately scaled to the mean to the extent of formlessness. War may be 'nine-tenths inactivity' but sudden exogenous events that characterise AWS battlecraft are now prone to be smoothed out of the unsupervised weapon's calculations. The coding practice here may be that the weapon continually computes new probabilities for its immediate world but the model is also based upon setting back to zero any inconsistent probabilities and then undertaking 'renormalizing' over the weapon's remaining possible outcomes. This coding routine is termed conditionalization whereby the weapon is calculating conditional probabilities for each set of possible causes and for each of its observable outcomes. Commanders must factor that AWS operation will be based upon a series of posterior probability distributions to use as its new prior in every next-time step. This then has the counter-intuitive effect of accelerating data obsolescence in the weapon's underlying datasets causing the AWS to be systemically instable. An example here is in AWS' movement and the fundamental requirement that all relevant navigable space must first be identified, then processed and then made 'map-ready' for each and every of the weapon's representations (its internal mapping of its state and immediate environment). The resulting dataset must dynamically be searched in real time to evaluate, first, available paths and then, second, the best path for the weapon after which the weapon's goals, values and action selection must all be revised to account for that newly selected path.

Coding issues have other command ramifications. The notion of meaning (here, the Cohort's intention) also creates challenge through the need to link sensed data to a particular representation on one of the AWS' learning planes. Each such linkage must be coded so that one associatively-connected entity can evoke another. In this way, the meaning of the commander's instruction is not restricted to just one association. The ability of the AWS to capture abstracts is similarly difficult. The information contained within a command must also be coupled to previously-given information as well as to information that is to follow. Research again highlights that this is difficult to define, difficult to test,

introduces process fragility as well as variability in that system's output. In this vein, commanders must note that nested structures and conditionals (which regularly characterize complex instructions) create similar syntactic issues. While it may be human practice to understand what has been directed without having to figure out exactly the meaning of the words, this does not translate across in machine coding. Nor can coding capture context (in this case, the information in adjacent instructions and from non-associated routines). Finally to this point, command and analysis both use several categories of facts within their syntax. Instances here might include indexical facts, normative facts, strong convictions, observations and hints, clarifications, reinforcements as well as basic ontological factual statements. All of these sub-types inform the human decision but must now be articulated only in AWS code. The challenge is also that such categorizations are volatile and change unpredictably according to new intelligence, new feedback as well as input, of course, from weapon sensors.

Anchoring issues and ambiguity in AWS deployment

The potentially deadly consequences of AWS' programming also highlight a need for reliable on-platform arbitration.[24] This reflects the weapon's lethality but also the Cohort's requirement to field a weapon system that is fit for purpose and which conforms to its expected deployment. Arbitration, however, cannot simply be based on most recent interactions. What, after all, might constitute a weapon's 'acceptable' delay? What for the local commander represents urgency? Code-based arbitration strategies and a remote means of settling decisions is challenging. An example is useful. Averaging protocols select weapon actions according to where the most conditions have been satisfied (also referred to as 'longest matching') but this should concern the Delivery Cohort as, by definition, they are approximations, reduce data precision and, as above, must compromise the platform's legal compliance.

This creates other trials for the local commander. The first relates to the *degree* of stepped change (the increment of learning termed 'anchoring'[25]) that each follow-on process exerts on the weapon's immediately prior set of beliefs. A second challenge is that weapon actions must comprise the appropriate reaction to *every* relevant sensed stimulus. AWS cannot offer erratic performance where only particular sensed inputs lead to weapon outputs. To this point, coding for weapon 'curiosity' might be a composite of the two states of 'novelty' and 'attraction' but each of these states may have a specific and often conflicting action routine in the weapon. Thus, programming for 'astonishment' might be the combination of, say, one third 'attraction' (go forward), one third 'withdrawal' (go backwards) and one third 'curiosity' (stay still and garner additional information). A third challenge then relates to AWS expression, for instance, of a delayed response, a measured response, a slight deferral or, more complicated, a variable response?

AWS methodology is then compromised by data ambiguity.[26] Communication is not just about dispatch and enactment of a message. There is typically a considerable gulf between the spoken or written word and an intended message. Here, lexical ambiguities might arise from omitted or imprecise script. Semantic ambiguities concern interpretative uncertainty while pragmatic ambiguity hobbles communicating parties with mildly different contextual bases. All of these challenges raise basic ethical issues for the commander and

Cohort. Are 'occasional mistakes' acceptable on the basis that unsupervised violence is similarly susceptible to ethical deficiency as human soldiers? And is a lower legal bar appropriate for machines with less lethality? The verso here is that there may be occasions when a hiatus offered by the compliant AWS (that neither needs to protect itself nor rush its action sequences) may prevent ethical transgressions that have characterised knee-jerk human behaviour in the chaos of battle. This ignores, however, difficulties arising from the simple firing sequence for such weapon instructions. Each routine (and the pattern created by those routines) may result in quite different outputs being triggered depending upon the order in which command instructions are processed by the unsupervised weapon. This requires arbitration through weighted average or 'centre of gravity' routines in an effort to optimise action selection but leaving unresolved the limitation in AWS' ML framework whereby symbols may only be read one at a time and, moreover, with each such symbol being processed on information collected from previous symbols. Indeed, algorithmic tools are demonstrably less useful in cases of growing ambiguity.

Uncertainty, of course, will be everywhere: The AWS may move erratically, enemy forces unexpectedly dissemble and battlefield obstacles move unpredictably. It is here that human leadership can leverage experience, judgement and intuition. It is also here where algorithms would appear to fall short of the human agent who is instead able to make difficult decisions in a fast, frugal and, crucially, an explainable manner.[27] At such higher levels of expertise, battlefield commanders do not even recognise that they are making decisions; rather, they are interacting with a changing situation and responding to patterns that they have long recognised. Training, experience, subjectivity and a wide grasp of context are all critical attributes that cannot currently (or foreseeably) be captured by code.

Leadership ramifications from AWS' action selection processes

In deciding to deploy weapons without human supervision, commanders also need to understand what will drive their weapons' actions.[28] AWS' sequences will be governed by what is termed an 'optimality notion', each unique to a weapon class and set by the Delivery Cohort as its initial set of decision rules. At each such step, the weapon is selecting an action with the highest expected utility. But various challenges arise from this utility model. The margin for error is very considerable and obviously comes with the possibility of unacceptably high-regret outcomes. As above, the weapon first run an internal computation on *all* possible actions in order to establish the pathway of highest utility, a considerable task given the almost limitless number of battlefield parameters including legal status and risks arising from poor execution. The utility function must also regulate what constitutes *appropriate* use of force, the involvement of colleague assets, consideration of next steps, a data audit ahead of an action sequence as well as post-event communication of each engagement. Commanders, of course, do this innately.

Similarly intractable is the Cohort's need to set goals governing weapon priorities and action selection. Leaders need to distinguish between values and goals in machine autonomy. Goals prompt an intelligent weapon to develop plans of action while values enable it to assess the comparative merits of such plans. If this process is stunted or inappropriately undertaken, the weapon will either be illegal or useless. Goals concern what must be undertaken at once, what should be undertaken next, the resumption of a task that

was previously discontinued and, more complex for the weapon, what actions should subsequently take place in order to capitalize on battlefield opportunities. Errors here may also have quite unforeseen battlefield consequences that include 'infrastructure profusion' where an independent weapon might unexpectedly allocate disproportionately large parts of its reachable resources into the service of some inappropriate internal goal.

What then is the expected model to manage the autonomous agent's goal setting and what are its leadership ramifications?[29] The difference between a weapon's current state and its desired states then becomes the weapon's observed error. The object of the AWS' action selection is thus to minimize that error. Two issues arise for leaders. The pace of this error correction is not obvious. It depends, for instance, on how often the error is computed and how much correction is then made on each feedback loop. Feedback loops, moreover, will be significantly less effective at modifying a weapon's higher-level action selection in particular in regard to prioritisation, coordination and collaboration. Second, the AWS must also be appropriately front-facing and determine its system state *ahead* of time. For this reason, independent weapon systems (and their deploying commanders) must be able to support parallelism, the complex ability to monitor and execute multiple actions at once, the risk being that sequential processing would otherwise risk missing events that might be critical to compliant and winning operation. Similarly, goals and behaviour models must also support a workable process of 'data forgetting' in order to mitigate information overflow or the retention of sub-optimal (or wrong) data sequences. Very little research is currently published on this issue. An unsupervised weapon, after all, should not generally forget acquired skills (termed 'catastrophic forgetting'), a well-recognised limitation of the neural network model. The issue, moreover, has several levels. How, for instance, should data age conflate with data redundancy?

The issue of AWS 'attention' relates to the focus of the unsupervised weapon.[30] The human brain appears to be free to choose what it looks at, listens to and thinks about. In benign conditions, human commanders can focus their attention as they please. This information access if not limited in any way, would lead in AWS to memory overflow and to what is termed in a machine as 'contradictory neural cacophony'. In order to be useful and compliant, the AWS must therefore select both the source and quantity of its information, what to process, what to store and which peripheral information should be attenuated into its decision processes, the so-called 'cocktail party effect'. Research confirms that this is not obvious. Similar to cognition, attention is also divided into the voluntary and involuntary.[31] The issue here again relates to arbitration and the need for an appropriate model to determine that one such sensor input be preferred over others. The issue also questions how the weapon's input intensities should be managed. After all, two variables that may be useless by themselves can be useful together. Similarly, a single variable that is useless by itself can then be instrumental with others. Nor can it be that the loudest signal is automatically the one upon which the weapon should focus. Both data sensitivity and data habituation will remain enduringly challenging to the deployment of AWS.

Implications arising from AWS' technical debt

The foregoing evidences the significant 'technical debt' that precludes the removal of human supervision from weapon operation. Indeed, its notion provides a useful metaphor

that anchors much of this paper. It links the consequences of poor system design in business to accumulating a 'financial debt', the assertion here being that such 'debt' is particularly prevalent in the removal of battlefield supervision. The causes of technical debt must drive the commander's allocation process. They arise from inappropriate architecture, from shortcuts resulting from commercial pressures, poor testing protocols and, more broadly, from poor whole-AWS understanding. They also therefore arise from an overall lack of project ownership, from poor technical leadership and the consequences of pervasive changes in weapon specification. Debt here is also a consequence of 'counterparty development' where a weapon's disparate software routines, once developed, must eventually be merged into a single source base. Scale compounds technical debt both through an exponentially growing number of interactions and the number of interdependencies required among developers and those charged with deploying these weapons. Commanders therefore need to realise that such debt compounds as projects evolve, including management of the weapon's configuration, its integration, its verification and validation (a catch-all expression for its testing) and, in the case of ML, determining its 'logical completeness'.

In discussing leadership ramifications, correction routines pose a further challenge given AWS are by definition independent and remote and will operate where remediation is correspondingly difficult. They also require that the correcting party factor both the specifics but also the distribution of any system error. Error cascades, moreover, are generally prone to deadlock whereby the local optimum for a learning system quickly becomes iterative so that neither the weapon component nor its attached routine can then be improved. This is also not helped by the weapon being a bundling together of disparate proprietary routines held together by 'glue' or 'spaghetti' code. Glue here relates to the quantity of supporting code that must be incorporated to allow data transfer within these routines. The phenomenon increases fragility, not least because glue code anchors the weapon to deployment idiosyncrasies and the initially programmed states of each autonomous component and so discourages experimentation, the intended essence of machine learning. The problem is well framed by research noting that machine learning systems end up being five per cent executive code and ninety-five per cent glue code.

Command decisions must also factor for the *configuration* of unsupervised weapons. Configuration routines are empirically less structures, less tested and a proven source of unpredictability.[32] There are several layers to this. First, configuration thresholds must anyway be overseen by responsible, experienced humans. Indeed, part of leadership is that this configuration process is well understood by the commander not least because of its inherent complexity but also because of frictions arising from in-field, ad hoc adjustment where soldier ingenuity and experience have traditionally been relevant. If a weapon updates on new data, then its old manually-set thresholds may be invalid, the more so given the weapons' likely different learning states. This also needs to be undertaken in real-time and then validated across the whole of the weapon system in order for that weapon to be both compliant and still valuable. The leadership challenge, moreover, is *what* metrics should be monitored given that a purpose of its ML is, of course, that the weapon adapts over time.

Finally, legal, social and political constraints will necessitate that such limits are set conservatively which, should an action limit unexpectedly trigger and the weapon close down, might compromise its operational usefulness. The issue here is that weapon sub-systems then become 'undeclared consumers', consuming the output of a particular prediction routine as input to another component of that sequence. As noted by UNIDIR, unintended feedback loops then form between weapon algorithms and the weapon's external world. Such loops are akin to filter bubbles in social networks whereby noise suppression mechanisms inadvertently suppress nonconforming data. For a leadership angle, the phenomenon actually makes it problematic to make *any* changes to the weapon's firmware. Indeed, as complexity grows and the number of critical components increases, commanders should assume mathematically that the probability of whole event failure increases. They must also factor that the total probability of accomplishing any task gets smaller with the addition of more objectives.

Leadership challenges arising from AWS' empirical operations

It is useful to consider operational rather than on-platform challenges arising from AWS deployment. An example here is the policy of target profiles that must comprise the backbone of operation for unsupervised weapons.[33] There are several *operational* constraints that should govern leaders' attitudes to using independent weaponry. Under adopted Rules of Engagement and customary LOAC, commanders must ensure that target profiles do not incorporate 'people representations' as a targeting filter in weapon engagement routines. Similarly, the embedded profiles against which sensed data is matched should not be based upon age, gender, race or other societal identities. Target profiles should not blend civilian and military objects and should not intermingle target objects unless civilians have first been removed from the area of potential engagement. All parties within the Cohort must also be confident that target profiles remain immutable and are not subject to the same learning that is intended for the balance of the weapon's systems. Indeed, there should be a positive obligation upon deploying commanders to understand the exact target profiles of each AWS systems deployed in a battleplan. A similar positive obligation should ensure that both spatial area and duration (over which an AWS' sensor-analysis-process can occur) remain under the control of the local human commander in order to fulfil legal obligations under LOAC. Finally to this point, a defined and proximate relationship should exist between the weapon's targeting process and the local responsible commander whose judgement is responsible and accountable. In this manner, the commander should have reasonable understanding about the actual force effects of each autonomous lethal engagement while imposing a limitation on the number of engagements that a system may undertake within an individual attack.

It is, however, the practical essentials arising from AWS deployment which present the most fundamental leadership challenges. Such tests will include the logistics of AWS' in-field replenishment (including the scheduling and advance planning of this function), their repair and servicing, polling AWS performance, on-going audit and accountability of these systems' actions as well as a seamless means of liaison between machine and command structure. Commanders must also understand each such weapon's capabilities and competences notwithstanding that each deployed system will, by definition, be more-than-fractionally different from its colleague weapon depending upon the learning path

individually encountered since its deployment. Other practical matters include the failure modes for such weaponry (outright veto, fail safe, fail dangerously and fail deadly) as well as the possibility of integrating otherwise independent assets into the commander's wider portfolio combat assets. A reliable abort mechanism is particularly problematic given the ramifications of cyber-attack, the Cohort's overarching requirement for a resilient and predictable asset against legal and ethical repercussions in instances of the weapon's catastrophic failure. Finally, the commander must own the communication channel to ensure both engagement and action parameters are dynamically set exactly as intended. This, of course, is difficult given the AWS will usually be operating in communications-denied environments out of the range that patches can upgrade and remediate system errors as well as being able to poll useful information about each's weapon's configuration, readiness and reliability.

Conclusion

Overview of AWS' higher-order and technical pitfalls confirms the enduring difficulties that follow from the models currently posited for removing weapon supervision. Two key leadership ramifications arise. First, commanders must at least retain fundamental scepticism around AWS' feasibility. The tipping point in the equation for the Cohort is to introduce autonomy in engagement sequences in assets that remain useful, predictable and compliant. The second and overarching ramification to commanders concerns context and whether wide-task, wide-capability autonomous weapons might ever be appropriate given the very many alternative levers available to leaders to achieve aims and objectives whether in-the-loop or just on-the-loop as opposed to accepting the likely chaos that will accompany independent software and hardware assets.

About the Author

Paddy Walker has degrees from Durham (BA Modern History), Cornell (MBA) and Buckingham (MA, Modern War Studies; PhD, *'Challenges to the Deployment of Autonomous Weapon Systems*'). He is co-chair of the London Committee of Human Rights Watch and is a Board member of weapons-NGO Article 36. J Leon Philanthropy Council supports campaigns arguing for a statutory instrument to ensure meaningful human control in lethal engagement.

About this article's sources

This article is derived in large part from the author's PhD thesis, *'Challenges to the Deployment of Autonomous Weapon Systems*'.[34] The author was previously commissioned in the British Army and the piece is also derived from his Autumn 2019 address to the Center for Army Leadership at RMAS Sandhurst.[35] Finally, its commentary on target profiles arises from work undertaken by Article 36, the weapons think tank NGO of which the author is a board member.[36]

Bibliography

Article 36, '*A structure for regulating 'autonomy' in weapon systems*', 25 November 2019, <http://www.article36.org/other-issues/definitions/target-profiles-as-a-basis-for-rules-relating-to-autonomy-in-weapons-systems/>
Blair Williams, 'Heuristics and Biases in Military Decision Making', *US Army Combined Arms Centre, Fort Leavenworth*, October 2010
Max Boot, *War Made New: Weapons and the Making of the Modern World* (USA:Gotham, Penguin, NY, 2006)
Center for Army Leadership, Library resources, https://www.army.mod.uk/who-we-are/our-schools-and-colleges/centre-for-army-leadership/leadership-speaker-series/
Christian Enemark, *Armed Drones and the Ethics of War: Military Virtue in a Post-Heroic Age* (Routledge, Oxford, 2014)
Joshua Rothman, 'The Art of Decision Making', *New Yorker,* 14 January 2019
US Field Manual, *101-5*, Chapter 5 ('The Military Decision-Making Process')
Paddy Walker, 'Challenges to the deployment of autonomous weapons', *Buckingham University*, PhD thesis, June 2019, see https://killerrobots.info/wp-content/uploads/2019/08/war-without-oversight-august-2019-phd-distribution-version.pdf
Blair Williams, 'Heuristics and Biases in Military Decision Making', *US Army Combined Arms Centre, Fort Leavenworth*, October 2010

Footnotes

[1] Paddy Walker, 'Challenges to the deployment of autonomous weapons', *Buckingham University*, June 2019, see https://killerrobots.info/wp-content/uploads/2019/08/war-without-oversight-august-2019-phd-distribution-version.pdf, p.5 ('*Abstract*'). Much of this paper is derived from the author's PhD thesis as referenced.
[2] Walker, 'Challenges', Chapter 10 ('*Oversight: Command and control constraints to AWS deployment*'), pp. 268-282.
[3] Walker, 'Challenges', Chapter 1 ('*Introduction*') and Chapter 3 ('*Drivers: Factors accelerating the removal of weapon supervision*'), p. 68 and generally.
[4] Academy Leadership, *Eleven Timeless Principles of Leadership'*, US Army Field Manual, 1951, generally.
[5] US Field Manual, *101-5*, Chapter 5 ('The Military Decision-Making Process'), p. 5-1.
[6] Walker, 'Challenges'
[7] Christian Enemark, *Armed Drones and the Ethics of War: Military Virtue in a Post-Heroic Age* (Routledge, Oxford, 2014) p. 99.
[8] Walker, 'Challenges', Chapter 4 ('*Deployment: Models for the removal of weapon supervision*'), pp. 91-98.
[9] Walker, 'Challenges', Chapter 4 ('*Deployment: Models for the removal of weapon supervision*'), pp. 108-123.
[10] Joshua Rothman, 'The Art of Decision Making', *New Yorker,* 14 January 2019, generally.
[11] Walker, 'Challenges', Chapters 6 ('*Wetware: Design challenges to AWS function*') and 7 ('*Firmware; Embedded process challenges to AWS function*'), generally.
[12] Walker, 'Challenges', Chapter 6 ('*Wetware: Design challenges to AWS function*'), pp.190 and generally, Chapter 8 (*'Software: Coding constraints to AWS function')* p. 238 and p. 282.
[13] Walker, 'Challenges', Chapter 6 ('*Wetware: Design challenges to AWS function*'), in particular section 6.6, *AWS Control Mechanisms*, p. 190 and generally.
[14] JSP 383, *Joint Service Manual of the Law of Armed Conflict*, 2004 Edition, DSDC (L) distribution, pp. 51-101 and generally.
[15] Walker, 'Challenges', Chapter 6 ('*Wetware: Design challenges to AWS function*'), pp.173, 186 and 208.
[16] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), pp. 213-230 and 232-236.
[17] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), pp. 213-249.
[18] Walker, 'Challenges', Chapter 6 ('*Wetware: Design challenges to AWS function*'), pp. 161-173.
[19] Walker, 'Challenges', Chapter 7, ('*Firmware: Embedded process challenges to AWS function*), generally.
[20] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), generally.
[21] Blair Williams, 'Heuristics and Biases in Military Decision Making', *US Army Combined Arms Centre, Fort Leavenworth*, October 2010, p. 7.
[22] Walker, 'Challenges', Chapter 11, ('*Conclusion*') in particular section 11.1 *The Nature of Deployment Challenges*, pp. 280-302.
[23] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), pp. 213-249.
[24] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), section 8.1 ('Coding methodologies'), p. 222.
[25] Walker, 'Challenges', Chapters 6 ('*Wetware: Design challenges to AWS function*') and 7 ('*Firmware; Embedded process challenges to AWS function*'), in particular section 8.5 ('Anchoring and goal setting issues'), pp. 236-240.
[26] Walker, 'Challenges', section 8.1 ('Coding methodologies'), pp. 219-229.
[27] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), in particular section 8.1 *Coding Methodologies*, pp. 219-229.
[28] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), sections 8.7 ('Action selection issues') and 8.8 ('Behaviour setting and coordination'), pp. 243-249.
[29] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), section 8.5 ('Anchoring and goal setting issues') and 8.6 ('Value setting issues'), pp. 236-243.
[30] ibid, pp. 240-246.
[31] Walker, 'Challenges', Chapter 8 ('*Software: Coding constraints to AWS function*'), in particular section 8.1 *Coding Methodologies*, pp. 219-229.
[32] Walker, 'Challenges', Chapter 2 (*Context: The role of context in removing weapon oversight*'), generally and Chapter 7 ('*Firmware: Design challenges to AWS function*'), specifically sections 7.1 ('Sources of Technical debt') and 7.2 ('Firmware ramifications of learning methodologies'), pp. 192-206.

[33] Article 36, '*A structure for regulating 'autonomy' in weapon systems*', 25 November 2019, <http://www.article36.org/other-issues/definitions/target-profiles-as-a-basis-for-rules-relating-to-autonomy-in-weapons-systems/> [accessed 12 February 2020], generally.
[34] Walker, 'Challenges to the deployment of autonomous weapons', *Buckingham University*, June 2019, see https://killerrobots.info/wp-content/uploads/2019/08/war-without-oversight-august-2019-phd-distribution-version.pdf, generally.
[35] Paddy Walker, 'Leadership Challenges to AWS Deployment', *Center for Army Leadership*, see https://killerrobots.info, Autumn 2019, generally. For CAL Sandhurst lecture, see https://www.youtube.com/watch?v=lYbZaZSMQA8.
[36] Article 36, *Target Profiles as a Basis for Rules relating to Autonomy in Weapon Systems*, <http://www.article36.org/other-issues/definitions/target-profiles-as-a-basis-for-rules-relating-to-autonomy-in-weapons-systems/>, 2 August 2019 [accessed 23 February 2020], generally.