International Review of the Red Cross (2015), **97** (900), 1313–1339.
The evolution of warfare
doi:10.1017/S1816383116000023

INTERNATIONAL REVIEW of the Red Cross

Factors shaping the legal implications of increasingly autonomous military systems

Tim McFarland*
Tim McFarland is a PhD candidate in the Asia Pacific Centre for Military Law at the Melbourne Law School. He is undertaking his doctoral studies as part of the research team working on the Emerging Technologies of Warfare as a Challenge to the Law of Armed Conflict project.

Abstract
This article identifies five factors that will shape the legal implications of the use of autonomous military systems. It contends that the systems which present legal challenges are those programmed to "make decisions" that are regulated by law. In so doing, they transfer control of, and responsibility for, those decisions away from those who have been traditionally seen as decision-makers to persons responsible for developing and deploying the system. The article also suggests that there may be limits to the extent to which the rules of international humanitarian law can appropriately regulate the altered relationship between soldiers and their increasingly autonomous weapon systems.

Keywords: IHL, autonomous weapon systems, weapon development, weapons law, military technology.

* The author would like to thank his supervisors Professor Tim McCormack and Dr Rain Liivoja, and Group Captain Ian Henderson, for their insightful comments on various drafts of this article. The research for this article was supported by the Australian Research Council's Discovery Projects funding scheme.

Introduction

It is well known that various States are actively developing military systems which will utilize advanced technologies to assist, supplement and, to an extent, replace human soldiers in combat roles. Development efforts underway today have already produced machines that can replicate some of the functions of fighter pilots[1] and sentries,[2] among others, and it appears inevitable that military system capabilities will continue to expand into areas traditionally the domain of human operators. These systems, commonly described as "autonomous",[3] promise vast operational changes in the conduct of hostilities over the next few decades. Accordingly, the prospect has sparked interest among lawyers working in fields related to armed conflict.

Initial legal analyses of the proposed systems have revealed two significant challenges facing those who attempt to reconcile these systems with international humanitarian law (IHL), beyond the usual challenge of obtaining reliable information about military developments. First, the subject matter is highly technical, with legal issues potentially arising from the nature of a host of technologies, most notably those relating to robotics and artificial intelligence. In addition to being specialized subjects well outside the typical field of expertise of lawyers, the state of the art in these areas is advancing rapidly and any detailed legal analysis is at risk of becoming obsolete before it is complete. Second, with few exceptions, systems currently in use appear to have a very low capacity for autonomous operation, generally below that needed to raise significant legal questions. Published plans give only general descriptions of the future forms of more advanced systems, the technologies driving them and the manner in which they will be used. Machines available for examination today, such as the Phalanx Close-In Weapon System[4] and missiles employing active terminal guidance and similar technologies,[5] display only the precursors to the capabilities which will be seen in future autonomous systems.

Due in no small way to the resulting uncertainty about the eventual capabilities of highly autonomous systems, there are many significant unanswered

1 Lockheed Martin, "UCLASS", available at: www.lockheedmartin.com.au/us/products/uclass.html (all internet references were accessed in May 2015).

2 Jean Kumagai, "A Robotic Sentry for Korea's Demilitarized Zone", *IEEE Spectrum*, 1 March 2007, available at: http://spectrum.ieee.org/robotics/military-robots/a-robotic-sentry-for-koreas-demilitarized-zone.

3 The concern here is not with remotely operated weapons such as the unmanned aerial vehicles (UAVs), or drones, currently being employed in various conflicts. Such devices are manually controlled by human operators in respect of their critical functions. While it is true that autonomous vehicles would also generally be crewless, this article discusses only issues arising from a machine's capacity for "making decisions" autonomously.

4 Raytheon, "Phalanx Close-In Weapon System (CIWS)", available at: www.raytheon.com/capabilities/products/phalanx/.

5 For an overview of these technologies see Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organization (NATO), *Precision Terminal Guidance for Munitions*, Advisory Report No. AGARD-AR-342, February 1997, available at: www.cso.nato.int/Pubs/rdp.asp?RDP=AGARD-AR-342.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

questions about their compatibility with existing IHL. It is important that those questions be answered in the near future; the expected magnitude of the changes in the conduct of hostilities demand at least a precautionary investigation, and their uncertain timeline presents a degree of urgency. Also, this is an all-too-uncommon opportunity for lawyers to prepare for the advent of a new military technology, or even influence development of new systems in advance of their being deployed, rather than attempt to address their misuse after the fact.

This paper sets out a basis for a systematic analysis of issues that may arise under IHL due to the use of autonomous military systems. No attempt is made here to conduct a full legal analysis, but only to present a discussion of some aspects of machine autonomy which are most likely to be significant, as well as some guidelines for identifying potential legal problems, in order to provide others with a basis for investigating each issue in detail. The essential characteristic of this approach is to focus the legal analysis on the decisions and actions of people who develop and work with autonomous systems, rather than on the operational capabilities of machines. Two terminological choices bear explaining.

First, while most discussion and most contentious legal questions relate specifically to autonomous weapons, this paper frequently refers more generally to "autonomous military systems". "System" in this context may refer to any piece of hardware or software, or any structured set of hardware and/or software components, that performs a defined task. A crewless aircraft or a weapon such as a gun turret is a system for the purposes of this paper, but so is a network of sensors distributed over a wide area, or a "cyber-weapon",[6] or even a piece of software that analyzes data to inform a commander's decisions. "System" may be considered roughly synonymous with "means of warfare"[7] as that phrase is used in various legal instruments.[8] This choice was made because, as discussed below, the technologies of interest are likely to be integrated into military hardware and software that would not normally be classified as weaponry but may still influence combat operations in a manner similar to autonomous capabilities in a weapon. For example, an autonomous intelligence, surveillance and reconnaissance (ISR) system that employs sensors mounted on UAVs, satellites, ships or other platforms to gather and process information about potential targets before providing it to a human weapon operator may play a similar role in relation to a firing decision as would an autonomous decision-making system which forms part of the weapon system itself.

6 A cyber-weapon is software and/or hardware "used, designed, or intended to be used" to conduct "a cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects": Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare,* Cambridge University Press, Cambridge, 2013, pp. 106 (Rule 30), 141 (Rule 41).

7 It is possible that use of an autonomous military system may also relate to a method of warfare, such as where the process leading to the decision to employ the system is at issue, or where the system has more than one mode of operation. For an example of the second case, see: *ibid.*, p. 142, paras 4 and 5.

8 See, e.g., Protocol Additional (I) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 1125 UNTS 3, 8 June 1977 (entered into force 7 December 1978), Part III.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

Second, in contrast to what is becoming common practice, this paper does not distinguish between "automated" and "autonomous" systems. Very generally, systems are elsewhere described as "automated" when they can perform only low-level tasks without outside assistance, or when they are locked into a rigid repetitive operational routine defined by fixed programmed procedures, and they are called "autonomous" when they can be given higher-level instructions and appear to exercise some degree of "choice" or "decision-making" ability in determining how to carry out those instructions.[9] This paper argues that such a distinction is artificial and poorly defined and is not useful in a legal analysis. The precise capabilities of individual systems do not fall clearly into distinct categories; rather, they appear to exist in a continuum of several dimensions, and it must be expected that as the relevant technologies develop further, system capabilities will likewise move beyond the bounds of current definitions. Lawyers wishing to discern durable legal principles should avoid framing them in terms of categories which are likely to change, and the approach presented here avoids that trap. Furthermore, the distinction relies on anthropomorphic notions of machines making decisions, exercising discretion and in some way doing something more than simply executing a program. While such terms may be useful metaphors for technical purposes, they are not accurate descriptions of the actual operation of the systems in question from a legal perspective, and are misleading if employed in a legal discussion. All systems which may be described as "automated" or "autonomous" are still merely machines, constructed, or programmed, to relieve a human operator of some decisions and actions that would otherwise have been carried out manually. This point is discussed in more detail below.

The key finding presented here is that lawyers should assess "autonomous" systems not according to whether they can act with some degree of independence or whether they display some human-like behaviour, but according to which decisions are delegated to them and how human operators relate to them. Systems that should be of interest to lawyers in this respect are those which relieve humans of decisions or actions that are regulated by law, and in so doing transfer some control of, and responsibility for, those decisions away from a manual operator to another party, perhaps the party who defines the behaviour of the system or the party responsible for its employment.

The paper contains three sections. The first discusses three aspects of machine autonomy which are relevant to a legal analysis of proposed military systems. The second discusses two aspects of the development proposals that have been put forward by military organizations. The third section explains how the proposed developments relate to IHL and offers guidelines for persons analyzing specific legal issues: to remember that responsibility for legally significant decisions remains with humans, to understand that the nature of those

9 There is, however, considerable disagreement about the precise definitions even in the technical community; see, e.g., the discussion in M. Shane Riza, *Killing Without Heart: Limits on Robotic Warfare in an Age of Persistent Conflict*, Potomac Books, Washington, DC, 2013, p. 13.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

decisions will change when working with autonomous systems, and to be careful to distinguish between legal and technical issues.

Aspects of machine autonomy relevant to a legal analysis

A rational assessment of the impact of making weapons "autonomous" must be based on a clear understanding of what machine autonomy actually is. In the context of a legal analysis, it is unnecessary to focus in great depth on the technical means by which autonomous behaviour will be achieved and much more important to accurately specify how autonomous capabilities will affect the interactions between such systems and the rest of the world, most notably interactions with operators and supervisors and those subject to the effects of the systems.

Unfortunately, while autonomy may initially seem unambiguous, it is difficult to fully capture the notion in a definition.[10] Indeed, the concept carries quite different meanings in different fields of study. Here it is used in a strictly technical sense, but even within the technical literature on machine autonomy it seems there are almost as many definitions as there are authors who define it. Nevertheless, at a high level we may say that most definitions focus broadly on two interrelated aspects of autonomy. Some definitions focus on how an operator interacts with the system: autonomous systems are those that can operate "without any form of external control for extended periods of time".[11] Others frame the issue in terms of the system's own capabilities: autonomy is "the capacity of a system to select and decide within limits its own behaviour".[12] Of course, these and other definitions each cover just a few of the facets of machine autonomy that are of interest for specific purposes, and lawyers wishing to conduct an investigation must also decide on which aspects are relevant for that purpose. Rather than provide yet another (inevitably incomplete) definition, this section attempts to orient the reader by presenting a high-level view of how machine autonomy is achieved in practice and then discusses the aspects of autonomous systems that are likely to be of greatest legal significance in a

10 See, e.g., the discussion in Henry Hexmoor, Christiano Castelfranchi and Rino Falcone, "A Prospectus on Agent Autonomy", in Henry Hexmoor, Christiano Castelfranchi and Rino Falcone (eds), *Agent Autonomy,* Kluwer, Boston, MA, 2003, p. 3.

11 George A. Bekey, *Autonomous Robots: From Biological Inspiration to Implementation and Control*, MIT Press, Cambridge, MA, 2005, p. 1. Similarly, "[a] system with a high level of autonomy is one that can be neglected for a long period of time without interaction": Michael A. Goodrich and Alan C. Schultz, "Human-Robot Interaction: A Survey", *Foundations and Trends in Human-Computer Interaction*, Vol. 1, No. 3, 2007, p. 217.

12 Charles François (ed.), *International Encyclopedia of Systems and Cybernetics*, Vol. 1, K. G. Saur, Munich, 2004, p. 51. Similarly, "[a]utonomy is a capability (or a set of capabilities) that enables a particular action of a system to be automatic or, within programmed boundaries, 'self-governing'": Defense Science Board, *The Role of Autonomy in DoD Systems*, US Department of Defense, July 2012, p. 1, available at: www.acq.osd.mil/dsb/reports/AutonomyReport.pdf.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

military context and which should be borne in mind by lawyers studying development proposals. In particular, as explained below, it is inappropriate to regard the process simply as weapon development.[13]

How autonomy is achieved

The analysis commences with a basic understanding of the operations of an autonomous weapon system. Designers generally represent autonomous systems as consisting of two main components: the "plant" is the system or process to be controlled, and the "controller" is the device which directly governs the behaviour of the plant.[14] The term "plant" is carried over from chemical engineering; in a military context it would refer to the equipment which, if not capable of autonomous operation, would be directly operated by a human, such as a vehicle or gun turret. The controller of an autonomous system consists of the hardware and software that manages the vehicle, weapon or other device according to a program provided by a developer.[15] Figures 1 and 2 give a conceptual outline of how these components work together in manually operated and autonomous systems. The solid arrows show the typical interactions between the various components. These diagrams do not relate to specific weapon systems; they merely describe the generic functionality of each type of system.

Several points can now be made about machine autonomy to inform a legal analysis.

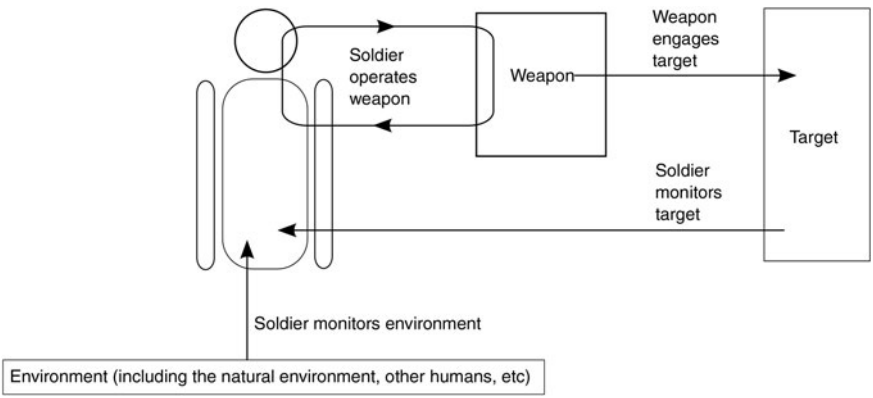

Figure 1. Manual weapon system.

13 The appropriateness of categorizing autonomous systems as weapons is also discussed in Hin-Yan Liu, "Categorization and Legality of Autonomous and Remote Weapons Systems", *International Review of the Red Cross*, Vol. 94, No. 886, 2012, p. 627.

14 See, e.g., Zdzislaw Bubnicki, *Modern Control Theory*, Springer, Berlin and New York, 2005, pp. 3–4.

15 For a general introduction to autonomous control systems, see Panos J. Antsaklis, Kevin M. Passino and S. J. Wang, "An Introduction to Autonomous Control Systems", *IEEE Control Systems*, Vol. 11, No. 4, 1991, p. 5.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

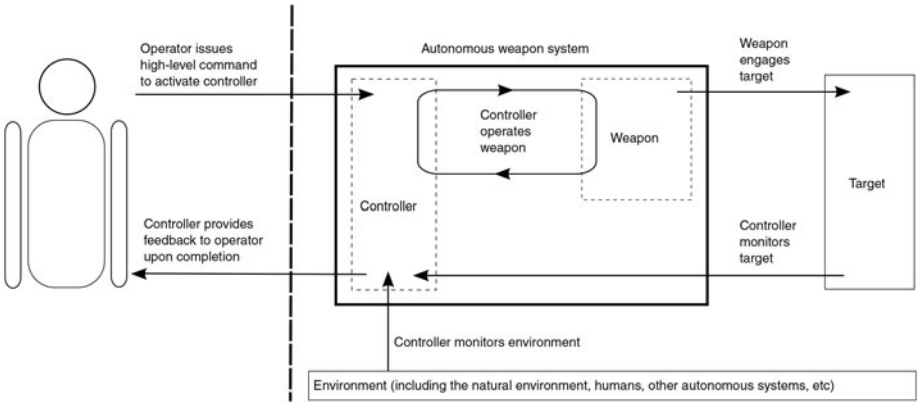

Figure 2. Autonomous weapon system.

Autonomy is neither fixed nor uniform

It is common for legal authors to refer to autonomous weapons as a discrete category of devices which are easily distinguishable from non-autonomous systems and to refer to the "level" of autonomy that a particular system exhibits as though such levels are intrinsic properties of particular systems.[16] This simplistic distinction does not reflect the range of capabilities of systems that exist today and it does not correspond with the development roadmaps[17] that have been made public by various armed forces. In fact, the levels of autonomy exhibited by proposed military systems may be expected to vary in complex ways.

Possible degrees of autonomy vary widely, as do the ways in which tasks are allocated between an operator and an autonomous system, and the behaviour of a system may be expected to change according to both the specific task being performed and the circumstances in which the system is operating. Establishing the relative degrees of control exercised by a human operator and a computer control system in respect of a particular action for legal or other purposes may be a complex process. The aim of the brief descriptions in this section is not to furnish the reader with all the technical knowledge necessary for such an undertaking, but to demonstrate the fluid nature of control over tasks in an environment where humans interact with advanced autonomous systems.

Degree of computer control

Machine autonomy is not an all-or-nothing capability; there is a continuum ranging from complete human control over some function to complete machine control.

16 See, e.g., Markus Wagner, "Taking Humans Out of the Loop: Implications for International Humanitarian Law", *Journal of Law, Information and Science*, Vol. 21, No. 2, 2011.

17 See, e.g., Army Capabilities Integration Center, US Army, *Robotics Strategy White Paper*, 19 March 2009, available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA496734; US Air Force, *Unmanned Aircraft Systems Flight Plan 2009–2047*, 18 May 2009, available at: www.fas.org/irp/program/collect/uas_2009.pdf.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

Many ways of describing this continuum have been proposed.[18] One of the best-known, which recognizes ten levels of automation, is reproduced in Table 1. Once again though, these "levels" are merely ways of describing points on a continuum; there are no discrete levels of machine autonomy in reality.

The US National Institute of Standards and Technology developed a somewhat more complex taxonomy, the Autonomy Levels for Unmanned Systems (ALFUS) framework.[20] The ALFUS framework characterizes autonomy levels as a three-dimensional model, in terms of mission complexity,

Table 1: *Sheridan and Verplank's ten levels of automation*

100% human control	1. Human does the whole job up to the point of turning it over to the computer to implement.
	2. Computer helps by determining the options.
	3. Computer helps determine options and suggests one, which human need not follow.
	4. Computer selects action and human may or may not do it.
	5. Computer selects action and implements it if human approves.
	6. Computer selects action, informs human in plenty of time to stop it.
	7. Computer does whole job and necessarily tells human what it did.
	8. Computer does whole job and tells human what it did only if human explicitly asks.
	9. Computer does whole job and tells human what it did only if it, the computer, decides[19] he should be told.
100% computer control	10. Computer does whole job if it decides it should be done, and if so tells human only if it decides he should be told.

Source: Thomas B. Sheridan and William L. Verplank, *Human and Computer Control of Undersea Teleoperators: Technical Report*, Massachusetts Institute of Technology, 1978, pp. 8–17.

18 Peter A. Hancock and Stephen F. Scallen, "Allocating Functions in Human–Machine Systems", in Robert R. Hoffman, Michael F. Sherrick and Joel S. Warm (eds), *Viewing Psychology as a Whole: The Integrative Science of William N Dember*, American Psychological Association, Washington, DC, 1998, p. 521.

19 "Decide" in this case does not imply a human-like decision-making capability; it simply means the action is initiated by the computer according to its programming rather than in response to a command from the human operator.

20 NIST Engineering Laboratory, National Institute of Standards and Technology, *Autonomy Levels for Unmanned Systems*, 16 June 2010, available at: www.nist.gov/el/isd/ks/autonomy_levels.cfm.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

environmental complexity and independence from a human operator, with each of those factors broken down further into a series of lower-level considerations.

Other organizations have proposed alternative taxonomies of autonomy levels intended for various purposes, such as NASA's eight-level model for its Spacecraft Mission Assessment and Re-planning Tool (SMART),[21] and the US Army Science Board's eleven-level model from its study on human–robot interface issues.[22]

It is not necessary to delve further into these rather detailed models. The important point to note is just that it is not useful to attempt to describe a machine as autonomous, semi-autonomous or manually operated as though those are discrete, objective categories. Autonomous capability is a continuously varying phenomenon.

Methods of human–machine interaction and task allocation

Even highly autonomous systems that may need no significant human input during completion of their allocated tasks will be operating alongside other entities, both human and electronic, and will need to exchange information with those entities in order to receive instructions, coordinate efforts, report results and so on. The study of human–computer interaction and more recently that of human–robot interaction are multidisciplinary fields which have received considerable and increasing attention over the last two decades as computing and robotic capabilities have expanded.[23] Several alternative paradigms have emerged to describe how this interaction may be structured and how tasks may be allocated to either entity, with implications for the roles of humans. Very broadly, humans may occupy either supervisory or collaborative roles when working with autonomous systems, and may move between those roles during execution of a series of tasks. For example, one taxonomy lists five roles that humans may play when working with robots: supervisor, operator, mechanic, peer and bystander.[24] These roles may not be static; one design paradigm, called mixed-initiative interaction, "refers to a flexible interaction strategy where each agent can contribute to the task what it does best".[25] In the mixed-initiative paradigm, the respective roles of human and computer (or robot) are often not determined in advance, but are "negotiated" in response to evolving circumstances. At different times either the human operator or the machine might have direct control over a task, with the other assisting, or they might be working independently.

21 Ryan W. Proud, Jeremy J. Hart and Richard B. Mrozinski, *Methods for Determining the Level of Autonomy to Design into a Human Spaceflight Vehicle: A Function Specific Approach*, NASA Johnson Space Center, September 2003, p. 4, available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA515467.

22 US Army Science Board, *Ad-hoc Study on Human Robot Interface Issues*, September 2002, p. 16, available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA411834.

23 For an overview, see M. Goodrich and A. Schultz, above note 11.

24 Jean Scholtz, "Theory and Evaluation of Human Robot Interactions", in *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, Hawaii, 6–9 January 2003.

25 Marti A. Hearst, "Trends & Controversies: Mixed-Initiative Interaction", *IEEE Intelligent Systems*, Vol. 14, No. 5, 1999, p. 14.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

Human–robot interaction is an active field of research, with new principles emerging continually. While it is safe to assume that humans collectively will continue to exercise a high level of control over robotic weapons employed in armed conflict, lawyers should be cautious when assessing the roles of, and the degree of control exercised by, particular individuals.

Variations by function

It is common to refer to autonomy as a property of a system as a whole, but in practice technologies which contribute to a capability for autonomous operation will be, and are being, applied to individual subsystems which form parts of military hardware and software. In respect of robotic systems, development is occurring separately in the areas of autonomous navigation, autonomous targeting and other functions required of an advanced weapon system. As systems with greater capabilities for autonomous operation are developed over the coming years, there is no reason to suppose that all their functions will be subject to the same degree of direct human supervision; it is probable that some functions, presumably those for which the cost/benefit analysis of automation is more favourable in some way, will be made "more autonomous" than others. A system may therefore be operating at more than one "level" of autonomy simultaneously, with respect to different tasks, and lawyers examining such systems will need to assess their behaviour, and the levels of human and machine involvement, in relation to particular functions of interest such as course planning, navigation or weapon release.

Looking more closely at the six-step targeting process used by the Australian Defence Force,[26] for example, weapon systems with the ability to autonomously locate and observe potential targets, and perhaps take some precautions to minimize collateral damage, may not be trusted to release a weapon autonomously, and may need input from a human operator or other autonomous systems to assess whether the target is a valid military objective and whether any expected collateral damage would be disproportionate to the anticipated military advantage of an attack.[27] If such a system were to be involved in an incident which led to an inquiry about whether a civilian might have been targeted, it would be important to establish at which stage an error may have occurred, and what levels of human and machine control were exercised at that stage, in order to determine whether a human operator exercised sufficient control to be held criminally liable. Similar considerations would also apply to a

26 Ian Henderson, *The Contemporary Law of Targeting*, Martinus Nijhoff, Boston, MA, and Leiden, 2009, p. 237.

27 For a more detailed discussion of how a balance between human control and autonomous operation may be achieved at different stages of the targeting process, see Mark Roorda, "NATO's Targeting Process: Ensuring Human Control Over and Lawful Use of "Autonomous" Weapons", Amsterdam Center for International Law Research Paper No. 2015-06, 2015, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2593697.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

"system of systems" situation involving communication between different systems with varying levels of autonomy.

Variations by circumstance

Whereas the functions of an autonomous system refer to tasks the system is performing, circumstances are the situations in which it is performing those tasks. As noted above, even where a designer has determined that a particular system function will be subject to a level of computer control, that level may vary according to circumstances, where the circumstances in question may be defined in terms of, for example:

- the phase of a mission, such as planning, initiation, implementation or termination;
- a disruptive event, such as where an unmanned aerial vehicle (UAV) that is ordinarily remotely operated may be programmed to autonomously return to base, fly in circles or even defend itself if it loses contact with the operator;
- an unexpected opportunity arising during a mission.

The US Department of Defense summarizes this variability thusly: "The key point is that humans and computer agents will interchange initiative and roles across mission phases and echelons to adapt to new events, disruptions and opportunities as situations evolve."[28]

Summary

Machine autonomy is a much more nuanced phenomenon than popular images of Terminator-like robots would lead one to believe. It is a rapidly advancing set of technologies, some still in their infancy, and many issues which will one day prove critical are no doubt yet to emerge, but two in particular stand out for lawyers investigating the area today. First, a machine's autonomous capabilities directly affect primarily its supervisors and operators, not necessarily (in the case of a weapon) those against whom it is directed. In particular, the role of the operator is altered in important ways but not eliminated; there is no such thing as "complete" autonomy in the sense of a machine operating entirely independently of any human. Second, the precise form of the relationship between operator and machine is likely to change across different systems, at different times, in relation to different tasks. Lawyers must be wary of drawing inferences purely on the basis of a weapon being described as "autonomous". One may gain a sense of the possible complexity of the relationships between an autonomous system and its operators and supervisors by examining the US Defense Science Board's recently proposed Autonomous Systems Reference Framework,[29] a system for allocating functions and responsibilities to either the

28 Defense Science Board, above note 12, p. 27.
29 *Ibid.*, p. 24.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

computer or one of several human operators or supervisors during the system design phase. This framework considers how autonomy supports each of the various human users of a particular system in a military hierarchy, how communication between those users may be facilitated, how allocation of tasks to human or machine may vary over the course of a mission, and several other factors.

Autonomy is about the relationship between machine and operator

In the context of robotics in general, and the proposals for autonomous military systems in particular, autonomy refers to a capability, not to a specific technology, a particular device or a certain behaviour. It is simply the ability of a system to perform its function, whatever that may be, with less interaction with a human operator than a manual system would require. Autonomy is thus concerned with the relationship between the system and its human operator, not the nature of the system's task or the manner in which it performs that task. As explained above, this relationship exists on a spectrum running from complete human control to (effectively) complete machine control over specific tasks, depending on the degree of autonomous capability present in the system.

When a manual system or process is replaced with a system that is capable of some degree of autonomous operation, the controller "steps into the shoes" of the human operator of the manual system to some extent. The operator's (or a system developer's) understanding of how to control the system is expressed in software and programmed into the controller. The physical means by which the operator manipulates the system is converted to a set of actuators[30] which the controller can activate. Some form of sensor or feedback mechanism is provided by which the controller can monitor the system. Other sensors may also be provided which allow the controller to monitor relevant environmental factors. The controller manipulates information from all these sensors in accordance with its programming and generates output signals which are sent to the actuators to regulate the system. This process is encapsulated in the well-known "sense-think-act" paradigm which is often used as the operational definition of a robot.[31]

Military development proposals often discuss autonomy in terms of the "observe, orient, decide, and act" (OODA) loop,[32] the model of a combatant's recurring decision-making cycle developed by US Air Force Colonel John Boyd.

30 An actuator is simply a device through which a controller controls a plant. One example would be an electric motor which pivots a gun turret based on a signal from the turret's software-based control system.

31 "Sense-think-act" refers to the continuous process by which a robot perceives its environment, uses that information to "make decisions" according to its programming, and acts on those decisions; see, e.g., G. A. Bekey, above note 11, p. 2.

32 See, e.g., US Air Force, above note 17, p. 16. The OODA loop may be seen as another expression of the "sense-think-act" loop, and some literature on human–machine interaction also refers to an OODA-like loop to describe four types of functions that may be performed by an autonomous system: information acquisition, information analysis, decision selection and action implementation. See, e.g., Raja Parasuraman, Thomas B. Sheridan and Christopher D. Wickens, "A Model for Types and Levels of Human Interaction with Automation", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 30, No. 3, 2000, p. 288.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

The OODA model describes the ongoing mental and physical processes involved in observing one's environment and responding to changes therein in pursuit of some goal. A high-level goal may involve several subtasks, each with its own OODA loop to be completed in pursuit of the overall goal. In a manual system, all steps in a loop are completed by a human: observing the environment to extract raw information, orienting oneself in relation to the environment by processing that information to form a useful model, making a decision based on that model and acting on the decision. In terms of the OODA loop, the purpose of developing autonomous systems is to assign part or all of the loop to a machine in order to realize some operational advantage such as greater speed or endurance, lower cost or less risk to the operator's life. A highly autonomous system is one that can execute most or all of the OODA loops required to achieve some goal, using only the operator's high-level instruction as guidance in the decision stage of a loop, such that the "nearest" human functions similarly to a commander. A system with a lower level of autonomy is only able to execute the lower-level loops, or only certain parts of a loop, and must work together with a human operator to achieve a high-level goal, casting that person as more of a collaborator. One example of such a system might be a radar warning receiver; a simpler system might only be able to detect a possible threat and issue a warning to a human operator, while a more highly autonomous system might be able to implement countermeasures, such as dispensing chaff and flares, without human intervention.

The fact that autonomy operates on the relationship between operator and machine carries three important implications for lawyers.

First, autonomous systems will perform some or all mission tasks in place of humans, but not necessarily differently to humans. There is nothing in the concept of machine autonomy itself that supports an inference that an autonomous system must necessarily perform a task in a different manner than would a human or team of humans performing the same task manually. This is not to say, of course, that future autonomous systems will function identically to an equivalent human-operated system; the "persistence" of systems that do not require constant human interaction, the ability to quickly integrate data from many sources and the ability to take greater risk than could a crewed system, among other things, will greatly enhance their performance. However, such differences, while very important operationally, are somewhat peripheral to the issue of autonomy. UAVs, for example, already allow for a high level of persistence without necessarily exhibiting any capabilities associated with a high level of autonomy and without raising the same legal issues. Such shared capabilities between remotely operated and autonomous systems frequently lead to confusion over the distinction between the two classes of machine, as both are capable of "standing in" for humans in different ways and both evoke images of machines fighting wars without a human presence. It is important for lawyers to keep the distinction in mind, though, as the two raise different legal issues. Today's UAVs and similar devices extend the physical reach and capabilities of human war fighters, as will autonomous systems, but autonomous systems will, in addition, alter the decision processes leading to the activation of a weapon.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

Accordingly, legal analysis of proposed autonomous systems must be based on the premise that the novel qualities of the system are likely to be found in the processes leading to activation of a weapon attached to the system. It is not useful to attempt to attribute a particular standard of performance to a system merely on the basis of it being described as having some capacity for autonomous operation; all that one can reliably say on that basis is that the human operator's direct involvement in part or all of the system's OODA loop(s) will be reduced or removed. The mere fact of reassigning a task from a human to a computer does not necessarily alter the performance of that task.

Second, if specified decisions and actions are to be transferred from human operators to autonomous systems, and those decisions and actions are the subject of legal obligations, then it is important to consider how those decisions and actions may be attributed to responsible persons, and whether those responsible persons are those specified in existing legal rules. Where legal obligations are borne by a human operator who is assisted by, or is working alongside, an autonomous system, one must ask whether that operator is still occupying the role contemplated by the law.

Third, it is incorrect to describe autonomous systems as being "independent"[33] machines that operate "without human control";[34] the relationship between human and machine is not severed, it is only modified. Choices made by hardware and software developers in the design stage will shape the behaviour of the systems from then on. Mission planners and others will also impose constraints on each mission; for example, in the case of an autonomous UAV, a flight plan must be filed in advance specifying the area to be covered and the duration of the mission (and the UAV must be made to adhere to that plan), while decisions about how much fuel and which weapons to carry will further guide and restrict what may happen during an operation.[35] In these ways, a human hand always provides some degree of guidance despite a possible lack of direct supervision.

Autonomous systems will still be machines

Lawyers must avoid falling into the trap of anthropomorphizing autonomous systems. It is uncontentious that today's sophisticated weapon systems are merely machines which execute instructions encoded in software, and it is argued here that future highly autonomous systems envisioned by designers will not be anything more than that. Certainly, they will be more complex and more capable in many ways; they will likely be able to utilize information from sources beyond the reach of today's systems and process that information in the face of more

33 M. Wagner, above note 16, p. 159.

34 Gary E. Marchant *et al.*, "International Governance of Autonomous Military Robots", *Columbia Science and Technology Law Review*, Vol. 12, 2011, p. 273.

35 William Boothby, "How Far Will the Law Allow Unmanned Targeting to Go?" in Dan Saxon (ed.), *International Humanitarian Law and the Changing Technology of War*, Martinus Nijhoff, Boston, MA, and Leiden, 2013, p. 56.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

uncertainty, and will function effectively in more chaotic and demanding environments. Such enhancements will, however, be due to improvements within the system's software and the hardware it controls; they will not fundamentally change the nature of the system such that it should be regarded as something more than a computer-controlled machine. The system's capabilities and limitations will still result, either directly or indirectly, from human decisions and actions.

The software-based controllers in today's automated weapon systems are essentially special-purpose computers running programs which control the weapon in place of a human operator. These controllers, although they may be highly specialized in design and purpose, are nevertheless forms of stored-program computer, a class of devices which also includes common items such as personal computers. The defining characteristic of stored-program computers is that instructions entered by a human programmer are stored in the machine's memory and drawn upon to govern its operation.[36] Barring a major technological shift, tomorrow's autonomous systems will employ essentially the same technology; the systems will still be controlled by software written by human developers.

The fact that even very complex programs are merely sets of predefined instructions is often obscured in discussions about sophisticated weapon systems, and indeed it is not always apparent to an observer that a complex machine is merely executing instructions rather than operating independently. Even systems with only low-level capabilities are often driven by programs with instructions of the form "if <X happens> then <do action A> else <do action B>", and this can make it appear that the system itself is "choosing" between two alternative courses of action, when in fact the choice was made in advance by the person who wrote the program; the expression of that person's will was merely waiting within the system's memory for the previously determined trigger to be detected. For example, if a hypothetical autonomous UAV has cameras and an image recognition program which matches people within its field of view against a database of pictures of known insurgents, an instruction like "if <camera image matches image in database with probability of more than 95%> then <aim and fire> else <keep searching>" would make it appear that the UAV itself is selecting targets, when actually the targets and the conditions under which they would be attacked were selected in advance by the system developers.

This reference to computing technology is included here because it is important that lawyers avoid being misled by references to "intelligent" machines having the capacity for "choice" or "truly autonomous" operation.[37] No computer is able to choose for itself whether or not to run a program stored in its memory, or to exercise discretion about whether or not to execute a particular instruction within a program; any such appearance of "choice" can only be the

36 William Aspray, "The Stored Program Concept", *IEEE Spectrum*, Vol. 27, No. 9, 1990, p. 51.
37 See, e.g., Chantal Grut, "The Challenge of Autonomous Lethal Robotics to International Humanitarian Law", *Journal of Conflict and Security Law*, Vol. 18, No. 1, 2013, p. 5.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

result of other instructions embedded in the software. Fundamentally, the only function of a computer is to run whatever software is installed on it.

Where a computer can be shown to be executing low-level express instructions encoded in software by a developer, it is easy to see that the computer is not acting independently in any legally significant way, but this may not be so apparent when a system with more advanced capabilities is "trained" to adopt some desired behaviour that relates to a significant action such as firing a weapon.[38] Similar challenges arise when human operators or commanders provide an autonomous system with only high-level mission goals, leaving the system to formulate a series of low-level subtasks. Of course, artificial intelligence and machine learning are complex fields that extend far beyond the scope of this paper,[39] but for the purpose of a legal analysis it is unnecessary, and even undesirable, to delve into the details of specific algorithms. Instead, lawyers should recognize that development of an artificially intelligent system is in fact just an exercise in software development, no different on the level of human activity from development of a simpler, low-level program. In both cases, the developer conceives of some desired behaviour for the system and writes a program intended to impart that behaviour to the system. The distinction is in the algorithms employed in the program: instead of directly encoding actions of the form "if <target matches these parameters> then <fire>", as would the developer of a simpler program, the developer of an "intelligent" machine writes a program the function of which is to formulate some optimum set of actions to be performed in response to environmental stimuli encountered during a mission. There is, in a sense, an extra layer of abstraction between the developer and the weapon firing, in that the specific sequence of events leading to discharge of the weapon may not have been in the developer's mind but may have originated in the data on which the system was trained; nonetheless, the end result of running the program is still a rule telling the system to fire a weapon, just as in a simpler program comprised of fixed rules. This extra layer of abstraction complicates the process of matching specific outcomes to specific commands from a human, but it does not change the fact that the computer is only executing instructions formulated by its developer.

To be clear, it is argued here that references to the ability of "intelligent" systems to make their own "decisions" are misleading. While such systems may not be programmed with precise predetermined responses to every situation they encounter, they are programmed with some system for developing a response and are thereby operating in accordance with their programming regardless of whether or not the specific behaviours they in fact adopt were or could have been

38 Training, in this context, means exposing an artificially intelligent system to sets of example data, representing the tasks it will be faced with and the correct responses, in an effort to induce behaviours which produce optimal outcomes at those tasks. Training is essentially inductive in nature, as the "real" situations encountered by a trained system will inevitably differ in some ways from the examples it was trained on, and is therefore error-prone.

39 For a general overview see, e.g., Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, NJ, 2009.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

foreseen during development or at the time of deployment. Such a machine is still just an instrument of the will of its developers and those responsible for employing it in some situation; it is not accurately characterized as an independent decision-maker. For example, UAVs with the ability to navigate autonomously may be able to respond to events during flight which are not specifically represented in their software, but they still do so according to programmed rules for responding to unexpected events. The behaviour originates not in the machines themselves, but in the minds of their developers.

Aspects of military development proposals relevant to a legal analysis

Autonomy will extend beyond weapon systems

As noted in the introduction, this paper discusses autonomous military systems, not just weapons. Even within an analysis that is focussed specifically on IHL, it is necessary to account for the effects of autonomous capabilities in systems which may not themselves perform any hostile act but may still have some impact on a decision or action that bears legal consequences.

The prime example of such a capability would be seen in an autonomous ISR system which locates, identifies and tracks potential targets. For example, the US Department of Defense recently announced its Autonomy Research Pilot Initiative (ARPI), which "seeks to promote the development of innovative, cross-cutting science and technology for autonomous systems able to meet future DOD system and mission requirements".[40] The ARPI invitation for proposals identifies ISR as one of its technical challenge areas:

> By increasing the level of machine perception, reasoning and intelligence on ISR platforms themselves, a more efficient workload balance can be achieved. This includes the management and closed loop control of ISR assets to adapt to their environments and mission circumstances to collect appropriate and relevant data.[41]

Preliminary moves in this direction are already occurring; in the United States, the Defense Advanced Research Projects Agency (DARPA) is engaged in the Military Imaging and Surveillance Technology (MIST) programme, which aims to "develop a fundamentally new optical ISR capability that can provide high-resolution 3-D images to locate and identify a target at much longer ranges than is possible with existing optical systems".[42] Systems developed under this

40 Department of Defense, *Autonomy Research Pilot Initiative (ARPI) Invitation for Proposals*, November 2012, p. 1, available at: www.auvac.org/uploads/publication_pdf/Autonomy%20Research%20Pilot%20Initiative.pdf.
41 *Ibid.*, p. 4.
42 See: www.darpa.mil/program/military-imaging-and-surveillance-technology.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

programme would be able to perform automated identification and recognition of potential targets.[43]

The motivation for pursuing this line of research as part of ARPI is that

> [t]oday's battlespace is creating an unprecedented increase in intelligence, surveillance, and reconnaissance (ISR) data. The [processing, exploitation and dissemination] analyst can become overwhelmed in trying to integrate and analyze these various data inputs (imagery, video, communication and human ISR data) while also trying to track targets, infer sources and provide analysis feedback (in real-time or post-analysis).[44]

It may be inferred that the increased "intelligence" of ISR platforms will be used to perform some processing of raw data before communicating the result to a human, thus placing the ISR system in a position of influence over a human operator's perception of the battlespace. Such processing of sensor data has obvious implications for the impact that an advanced ISR system may have on a decision to fire a weapon or perform some other action which carries significant legal consequences.

Somewhat further removed causally from combat operations, but still with some peripheral influence on outcomes, are autonomous vehicles which will carry cargo, supplies and even people. There are many examples of efforts being made to automate resupply operations, such as the US Office of Naval Research's Autonomous Aerial Cargo/Utility System (AACUS) programme.[45] Following on from the success of the K-MAX unmanned cargo resupply helicopter[46] in Afghanistan, the aim of the AACUS programme is "the development of advanced autonomous capabilities to enable rapid cargo delivery by unmanned and potentially optionally-manned Vertical Take Off and Landing (VTOL) systems".[47] The programme will produce a system which can be installed in suitable aircraft to respond to calls from deployed units, autonomously (but under some supervision) planning its route, avoiding obstacles and bad weather and choosing a suitable landing site to deliver supplies and, eventually, evacuate casualties.[48] While vehicles such as those controlled by AACUS-like systems will not carry weapons, they will still be large, fast-moving objects that must operate in proximity to people and other vehicles and may carry hazardous materials or

43 Strategic Technology Office, DARPA, "Military Imaging and Surveillance Technology – Long Range (MIST-LR) Phase 2", Broad Agency Announcement No. DARPA-BAA-13-27, 12 March 2013, p. 6, available at: www.fbo.gov/index?s=opportunity&mode=form&id=78b0ddbf382678fa9ace985380108f89&tab=core&_cview=0.

44 Department of Defense, above note 40, p. 4.

45 Office of Naval Research, "Autonomous Aerial Cargo/Utility System Program", 27 September 2013, available at: www.onr.navy.mil/en/Science-Technology/Departments/Code-35/All-Programs/aerospace-research-351/Autonomous-Aerial-Cargo-Utility-AACUS.aspx.

46 Lockheed Martin Corporation, "K-MAX", available at: www.lockheedmartin.com/us/products/kmax.html.

47 Mary Cummings and Angelo Collins, *Autonomous Aerial Cargo/Utility System (AACUS): Concept of Operations*, Office of Naval Research, p. 2, available at: www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2012/12-004-CONOPS.ashx.

48 *Ibid.*

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

present other dangers.[49] They therefore introduce some legal risk, for example, if used to carry wounded soldiers.

The presence of autonomous capabilities in systems ancillary to combat operations will therefore necessitate a broader view of autonomy than simply as a property of a new type of weapon. In addition, as discussed in the next section, the ability of the AACUS and similar systems to negotiate directly with other unmanned systems[50] may complicate the task of even identifying which systems are relevant to a particular investigation.

Autonomous systems will collaborate with each other

As discussed above, autonomous systems alter the relationship between weapon and operator. Of similar importance is the relationship between nominally separate autonomous systems. Most development roadmaps published by military and government organizations make it clear that autonomous systems must collaborate with each other.[51] This interoperability may take a range of forms.

Many current research projects, both military and civilian, focus on the behaviour of groups of decentralized cooperating robots, colloquially known as "swarms", which work together as a single system in pursuit of some goal. One example of this is the US Army's Micro Autonomous Systems and Technology (MAST) programme,[52] which aims to create "systems of diverse autonomous mobility platforms equipped with miniature sensors to quickly, quietly, and reliably explore and find targets";[53] that is, teams of small air and ground vehicles which will be used by soldiers to explore and map complex environments such as urban settings or caves. The possibility of equipping swarms with weapons has also been raised.[54]

A somewhat different form of collaboration between autonomous systems is employed in development projects such as DARPA's Hydra programme.[55] The Hydra programme "aims to develop a distributed undersea network of unmanned payloads and platforms",[56] essentially a system of unmanned submersible platforms which would be used to launch a variety of UAVs and unmanned undersea vehicles (UUVs) close to enemy operations. Although still in its very early stages, this programme envisions a system in which the submersible

49 Anthony Finn and Steve Scheding, *Developments and Challenges for Autonomous Unmanned Vehicles: A Compendium*, Springer, Berlin, 2010, p. 156.
50 M. Cummings and A. Collins, above note 47, p. 2.
51 See, e.g., US Department of Defense, *Unmanned Systems Integrated Roadmap FY2011-2036*, No. 11-S-3613, 2011, pp. 49–50.
52 US Army Research Laboratory, "Micro Autonomous Systems and Technology (MAST)", 25 February 2011, available at: www.arl.army.mil/www/default.cfm?page=332.
53 MAST, "Research Thrusts", available at: www.mast-cta.org.
54 Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*, Penguin Press, New York, 2009, p. 229.
55 Federal Business Opportunities, "Hydra", 22 August 2013, available at: www.fbo.gov/index?s=opportunity&mode=form&id=4cc32f06144bd6f3eba18655135d6155&tab=core&_cview=1.
56 DARPA Tactical Technology Office, "Hydra", available at: www.darpa.mil/Our_Work/TTO/Programs/Hydra.aspx.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

"mothership" interacts directly with its UAV and UUV payloads; for example, UUVs would dock with and recharge from the mothership, collect intelligence information for use on their mission, and afterwards transfer information acquired on the mission back to the mothership to be sent on to authorities.[57]

These and similar projects, and the tactics and strategies that will be associated with them, are part of a broad general trend toward more closely interconnecting formerly disparate components of a military force. Perhaps the most widely known expression of this trend is the doctrine of network-centric warfare (NCW),[58] an approach to conducting hostilities which emphasizes the importance of sharing information between force elements in order to best utilize that information and the force's capabilities. It was originally developed in the United States, but similar approaches to interconnecting military systems are being pursued, or at least have been proposed, by NATO[59] as well as in Australia,[60] the United Kingdom[61] and other countries. One of the expected benefits of such information-sharing is the possibility of a degree of decentralized operation, or "self-synchronization", wherein two or more entities can interact directly and coordinate their efforts without employing a traditional hierarchical command and control structure.[62] Relevantly, where some of those integrated systems have significant levels of autonomous operation, the sharing of information will not necessarily be conducted or directly supervised by a human, so that, for example, if intelligence gathered by one autonomous system is utilized in an aggressive action by another autonomous system, all the systems involved may become part of a legal investigation.

While concepts such as NCW are not intrinsically linked to development of autonomous capabilities in military systems, they offer visions of the type of environment in which such advanced systems are likely to operate, and of the degree of integration between military systems that may become the norm. Lawyers should be aware that if this trend continues as appears to be planned, it will become increasingly difficult to separate one autonomous system from another in respect of involvement in some incident. Such a development would

57 John Keller, "DARPA Considers Unmanned Submersible Mothership Designed to Deploy UAVs and UUVs", *Military & Aerospace Electronics*, 23 July 2013, available at: www.militaryaerospace.com/articles/2013/07/darpa-uuv-mothership.html.

58 See, generally, David S. Alberts, John J. Garstka and Frederick P. Stein, *Network Centric Warfare*, 2nd revised ed., Department of Defense Command and Control Research Program, 1999.

59 NATO, "NATO Network Enabled Capability", 27 October 2010, available at: www.nato.int/cps/de/SID-815535E4-57782C82/natolive/topics_54644.htm.

60 M. P. Fewell and Mark G. Hazen, *Network-Centric Warfare: Its Nature and Modelling*, Defence Science and Technology Organisation, September 2003, available at: http://dspace.dsto.defence.gov.au/dspace/bitstream/1947/3310/1/DSTO-RR-0262%20PR.pdf.

61 UK Ministry of Defence, *Network Enabled Capability*, JSP 777, 2005, available at: http://webarchive.nationalarchives.gov.uk/20121026065214/http://www.mod.uk/NR/rdonlyres/E1403E7F-96FA-4550-AE14-4C7FF610FE3E/0/nec_jsp777.pdf.

62 D. S. Alberts, J. J. Garstka and F. P. Stein, above note 58, p. 175; for a fuller discussion of the status of self-synchronization, see B. J. A. van Bezooijen, P. J. M. D. Essens and A. L. W. Vogelaar, "Military Self-Synchronization: An Exploration of the Concept", in *Proceedings of the 11th International Command and Control Research and Technology Symposium*, Cambridge, 26–28 September 2006, available at: www.dodccrp.org/events/11th_ICCRTS/html/papers/065.pdf.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

potentially complicate the processes of reviewing new weapon systems, and of determining liability of persons involved in developing or operating those systems. Indeed, attempts to draw distinctions between separate systems, such as between lethal and non-lethal systems, may become increasingly artificial. Each device that forms part of a "system of systems" will exhibit both individual behaviour as a system in itself and group behaviour as a component in a larger network.

Autonomy and IHL

Legal effect of employing autonomous military systems

Despite the scale of the impending changes to the composition of armed forces[63] and the practice of conflict, one may still ask why the prospect of autonomous weapons being used in armed conflicts is of particular legal interest. Innovative machines of widely varying levels of sophistication, many of them far too complex for a non-specialist to understand in detail, have been utilized in combat for millennia, and rarely has there been much difficulty in applying existing principles of law to their use. Indeed, systems, including weapons, with a more limited capacity for autonomous operation are already being used in combat without any serious contention that their use is illegal. So why might a more highly autonomous weapon raise significant legal issues? That is, why might use of an autonomous weapon affect a State's ability to meet its obligations under IHL, or affect a court's ability to adjudicate possible violations of that law?

The answer, and the key legal distinction between autonomous systems and other complex military systems, is that machine autonomy affects the process of deciding to perform an action, whereas other complex systems have an effect only after a decision has been made.[64] This does not mean, as has been suggested elsewhere,[65] that sufficiently autonomous systems will themselves make decisions. Rather, the two critical points, explained below, are (i) that certain decisions which are currently made in the course of an armed conflict would be transferred away from traditional decision-makers and would instead be made, in effect, by the people who define the behaviour of the autonomous system and the people responsible for employing it; and (ii) that in so doing, the character of those decisions would be altered.

63 Sharon Gaudin, "U.S. Military May Have 10 Robots per Soldier by 2023", *Computerworld*, 14 November 2013, available at: www.computerworld.com/s/article/9244060/U.S._military_may_have_10_robots_per_soldier_by_2023.

64 It is understood that this point may raise further questions about matters such as what constitutes a "decision" for the purposes of legal analysis, and what conditions are required for a decision to be considered to be within the immediate control of an autonomous system.

65 Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons*, Ashgate, Farnham, 2009, p. 33.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

Autonomy reassigns operational decisions

An autonomous machine runs software which approximates the process of deciding to perform an action, or some part of that decision, followed by other software which directs the machine to perform the action. In other words, the person who would otherwise have had to decide whether to perform some act is partly or fully relieved of that decision, or the person who would otherwise have been in a position to initiate or prevent the commission of an act is partly or fully removed from such a position. Instead, some part of the decision is effectively embodied in software written by people who would not normally have meaningfully contributed to the decision to perform the resulting act. Several novel legal questions then arise. For example, must autonomous weapon systems be designed such that those responsible for planning an attack exercise the same control that they could exercise over human soldiers? Is it possible that weapon developers, or other persons who help define the behaviour of an autonomous weapon system, may be criminally liable for violations of IHL? To what extent is it legal to automate a function, such as a proportionality analysis, when the law places responsibility for that function on humans?

When working with systems that currently are operating at lower levels of autonomy, operators are relieved of only low-level decisions that do not carry significant legal consequences in themselves. For example, a "fire and forget" missile which autonomously evades interception en route to its target does not thereby relieve the operator who pressed the "fire" button of any legally significant decision. When a system operates with a greater degree of autonomy, however, there is potential for the machine's control system to emulate more significant decisions. Where an autonomous ISR system or a targeting system connected to a weapon plays some substantial part in selecting a target or making a recommendation to fire (one example being the Patriot missile system's ground-based radar),[66] the human weapon operator, if there is someone who can reasonably be considered as such, is no longer performing all parts of the mental process leading to firing on that target and is no longer the sole or even principal decision-maker providing input into a firing decision. The machine itself cannot be considered a decision-maker for legal purposes,[67] so the decision is instead partly or fully attributable to the people responsible for the behaviour of the autonomous system (bearing in mind that identifying those individuals and organizations may be a complex endeavour in itself), and by the people responsible for the decision to employ the weapon.

Autonomy changes the character of decisions

Transferring operational decisions away from the people who have traditionally made them to those who define the behaviour of autonomous systems necessarily alters the character of those decisions in three interrelated ways.

66 Raytheon, "Patriot", available at: www.raytheon.com/capabilities/products/patriot/.

67 That is, machines are not subjects of IHL and, it is suggested, the possibility that they may become so in the future is remote.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

First, the generality of the decisions necessarily increases. When human decisions made "on the spot" in respect of specific situations are replaced with, or supplemented by, programmatic instructions that had been previously provided to a machine, decisions about individual acts in specific situations are replaced with broader policy-like choices applicable to the range of situations that match whatever parameters had been previously provided to the machine.

Second, the timing of the decisions changes. Decisions about whether and how to perform an action via an autonomous system are effectively made at the time the relevant behaviour is programmed into the machine and at the time the decision is made to employ the weapon, rather than at the time the situation arises in a conflict. This is important because, firstly, it may have implications for the temporal scope of application of IHL,[68] and secondly, it requires an assumption that situations in which the machine is deployed in the future will not differ in important respects from those envisioned at the time the machine was developed and tested.

Third, the informational basis on which the decisions are made is altered. Decisions implemented via an autonomous system cannot be based on direct (or even indirect) observation of the situation to which the decision relates; rather, they must be based on whatever information is available through experience and foresight at the time the machine is programmed, and then confirmed when the decision is made to employ the weapon.

One result of these changes is that the causal link between a specific human decision and a specific action (or repeated action) or a specific outcome, such as a particular target being fired upon, may be weakened when the decision is enacted partly or fully via an autonomous system.

The above changes in the character of decisions occur whether actions are explicitly programmed into a machine, or whether technologies of artificial intelligence are employed to allow the machine to adapt its behaviour dynamically; in either case the developers of the machine, or those they answer to, exercise control over the behaviour of the system and exert that control by defining a goal and equipping the system with some means to achieve that goal.

Identifying legal issues

Such an alteration in military decision-making is important because IHL also operates in relation to decisions. In seeking to "limit the effects of armed conflicts for humanitarian reasons",[69] the law attempts to guide the decisions and resulting actions of persons individually and parties to conflicts collectively. Legal objections to the use of machine autonomy in armed conflict are most likely to

68 For a discussion of one such question, see Tim McFarland and Tim McCormack, "Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes?", *U.S. Naval War College International Law Studies*, Vol. 90, 2014, p. 361, available at: www.usnwc.edu/getattachment/ed8e80ad-b622-4fad-9a36-9bedd71afebe/Mind-the-Gap–Can-Developers-of-Autonomous-Weapons.aspx.

69 ICRC, "War and International Humanitarian Law", Geneva, 2010, available at: www.icrc.org/eng/war-and-law/overview-war-and-law.htm.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

be found where the technologies of autonomy and IHL operate on the same decision; that is, where use of a weapon or other system with a capacity for autonomous operation alters a decision-making process, or a decision outcome, beyond what is permissible under IHL or otherwise hinders the operation of the law. Given the low levels of autonomous capability present in today's weapon systems, such an event is still hypothetical, but plausible scenarios may be postulated: perhaps if an autonomous ISR system were to provide information directly to an autonomous weapon system such that human commanders could not independently verify its accuracy before activating a weapon, it might be seen as a violation of the obligation to take precautions in relation to an attack.

This may seem a vague criterion, but given the expected range of forms and applications of autonomous systems and the current early stage of development, it is perhaps the most that can be reliably said. The following guidelines may, however, be useful in identifying legal issues.

Legal decision-makers will continue to be human

IHL is anthropocentric by nature and design. With one of its primary purposes being "to soften … the evils of warfare, to suppress its useless hardships and improve the fate of wounded soldiers on the field of battle",[70] its focus is necessarily on people: those wielding weapons and those subject to the effects of those weapons. Developments in the law are driven by the need to regulate and account for human decisions and actions, and the need to relieve unnecessary human suffering.

This is perhaps a trite reminder, but it is useful in this context. The growth of a new technology, especially one considered by some to herald a "revolution in military affairs",[71] naturally draws attention to the details of the technology and the possibilities it presents. It is often tempting to treat the technology (or the devices which employ it) as the focal point of law, rather than returning one's attention to people. In some cases that is appropriate; where a technology or device is expected to have a narrow and well-defined set of effects on the conduct of hostilities and on the people involved, it may be expedient to treat that technology or device as embodying a particular effect on the law. Laser weapons intended to permanently blind a person, for example, have a narrow and well-defined effect which can be readily analyzed for compatibility with the laws of armed conflict and can be readily seen to necessarily exceed the bounds of those laws; as such, it is appropriate for lawyers to focus on those devices as representing a certain undesirable effect and to regulate them accordingly.[72] However, that is not the case with autonomous systems.

70 Convention for the Amelioration of the Condition of the Wounded in Armies in the Field, 22 August 1864 (entered into force 22 June 1865), Preamble.

71 P. W. Singer, above note 54, p. 203.

72 Additional Protocol (IV) to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

The addition of autonomous capabilities to existing and new military systems presents a different challenge. A system's capacity for some degree of autonomous operation does not map directly to a readily determined effect that may be assessed against the requirements of IHL. As discussed above, an autonomous control system directly affects only the relationship between weapon and operator, and thereby the behaviour of the operator, not necessarily the behaviour of the weapon being controlled. Also, given the variability in time and function of what may be classified as machine autonomy, it is difficult to make general statements about an operator's or developer's behaviour (thoughts and actions) in relation to a particular event on the battlefield. The most that can be said is that lawyers must focus their attention on that behaviour in determining where legal issues may arise. Technologies giving rise to a capacity for autonomous machine operation merely inform those determinations.

Distinguish between technical and legal issues

Some of the issues most commonly raised by legal authors are essentially concerns about whether a proposed autonomous system will meet some standard of performance that is deemed necessary to comply with the law. Such concerns have so far most often related to the legal principles of distinction and proportionality.[73]

Regarding distinction, for example, several authors have argued that no existing robotic system can reliably distinguish between a combatant and a civilian, a shortcoming which is variously attributed to limitations in the effectiveness of sensor systems,[74] to the inability of controller software to "understand" context and human behaviour sufficiently well, and to the difficulty of sorting relevant from irrelevant information in complex situations,[75] among other factors. Similar arguments are used to show that robots cannot judge whether a proposed attack would satisfy proportionality requirements, with authors most often pointing to the qualitative and subjective nature of the problem.[76]

Other authors point to more general limitations in the ability of robotic systems to function reliably in conflict. Asaro cites the "highly limited capabilities for learning and adaptation" of autonomous systems as a reason why "it will be difficult or impossible to design systems capable of dealing with the fog and

Effects (Protocol on Blinding Laser Weapons), 1380 UNTS 370, 13 October 1995 (entered into force 30 July 1998).

73 See, e.g., C. Grut, above note 37; Peter Asaro, "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making", *International Review of the Red Cross*, Vol. 94, No. 886, 2012, p. 687; M. Wagner, above note 16.

74 See, e.g., Benjamin Kastan, "Autonomous Weapons Systems: A Coming Legal 'Singularity'?", *Journal of Law, Technology and Policy*, No. 1, 2013, p. 60.

75 M. S. Riza, above note 9, pp. 129–132.

76 See, e.g., Noel E. Sharkey, "The Evitability of Autonomous Robot Warfare", *International Review of the Red Cross*, Vol. 94, No. 886, 2012, p. 789.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

friction of war".[77] Others note the difficulty of reliably predicting the behaviour of complex autonomous systems.[78]

No doubt there is considerable strength in many of these arguments, given the state of the relevant technologies today. However, when an objection to the use of autonomous systems is presented as a statement about the ability of such a system to perform at a standard required by IHL in a given situation, that statement, even if accurate, relates only to a specific set of machine capabilities and so is only valid at a certain point in time; the objection may fall away as technology moves on. Lawyers should therefore be wary of making general statements about the legality of autonomous systems on the basis of arguments of this form. Such an objection would serve as the basis for a legal ruling (perhaps as guidance for a State's weapons review process) only if it could be expected to endure, such as may be the case if a factor could be identified which appears to positively prevent systems with some capacity for autonomous operation from ever performing in compliance with IHL, or if something inherent in the nature of machine autonomy would violate IHL. Examples from outside the realm of machine autonomy would include the bans on perfidy (due to the nature of the act) and the use of sarin gas (due to its inherently indiscriminate effects). If no such factors can be identified, it would seem more appropriate to state merely that a particular system could or could not be used legally in a specific situation. For this reason it is suggested that investigations into the legality of autonomous systems should include a qualifying question: if the system could, even hypothetically, be made to function in a given situation as effectively as a non-autonomous system (whether human or otherwise), would there be any legal basis for objecting to its use? If the answer is clearly "no", then the objection is more properly seen as a technical challenge that must be overcome before the system can be legally employed, or perhaps as justification for limiting the situations in which the system can be used until that challenge is overcome. For example, if an autonomous targeting system could be made to distinguish combatants from civilians as reliably as, or more reliably than, human soldiers do unaided, would it be proper to object to its use? It is suggested that objections to the use of autonomous systems that refer to the difficulty of predicting their behaviour in all circumstances, or to the possibility of computer errors and malfunctions occurring after activation, would also fall within this category. The possibility of malfunctions, in particular, is not unique to autonomous systems; it is, rather, a concern that must be addressed in relation to any military system that relies on a computer for some part of its operation.

If, however, the answer is "yes" (that is, regardless of how effective an autonomous system is, its use does not currently fit within the constraints of IHL), then lawyers need to look more closely to determine how development of

77 P. Asaro, above note 73, p. 692.

78 A. Finn and S. Scheding, above note 49, p. 36: "As the degree of autonomy increases, so it becomes increasingly difficult to predict the sum state of the system." Also see the discussion of developer accountability on p. 183.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press

the autonomous system should be limited, or guided, or how the law should be developed. This may be the case if, for example, it is found that the law requires some degree of direct human control over each individual attack or act of violence; if that is so, it would be impermissible to place an attack entirely under the control of an autonomous system, regardless of how well the system performs.

Conclusion

Even at this early stage of development, it appears likely that the use of sufficiently autonomous military systems may test the limits of IHL in a range of fundamental ways. The novel quality of autonomous systems is that they reassign operational decisions and, in so doing, change the nature of those decisions. In the case of armed conflict, situational decisions made by individual combatants would be replaced with more general choices made by people who define the behaviour of the autonomous systems.

The expected variability of the degree of autonomous operation in proposed systems, the likely range of applications and the possibility of autonomous systems collaborating directly with each other combine to greatly complicate analytical approaches which focus on the behaviour of specific machines in specific situations. Instead, it is recommended that lawyers treat machine capabilities as the result of decisions made by weapon developers, and focus attention on the roles played by those developers and by operational personnel who deploy and work with autonomous systems. In particular, the novel legal consequences of autonomous capabilities relate not to how well a machine performs its function, but only to how its developers and operators, if any, are involved in the outcome.

https://doi.org/10.1017/S1816383116000023 Published online by Cambridge University Press