Springer

Dear Author

Here are the proofs of your article.

- You can submit your corrections **online** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- Please return your proof together with the permission to publish confirmation.
- For **fax** submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the journal title, article number, and your name when sending your response via e-mail, fax or regular mail.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.

Please note

Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL:

`http://dx.doi.org/10.1007/s43154-020-00024-3`

If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information, go to: http://www.springerlink.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us, if you would like to have these documents returned.

The **printed version** will follow in a forthcoming issue.

AUTHOR'S PROOF

Metadata of the article that will be visualized in OnlineFirst

1	Article Title	**Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues**	
2	Article Sub- Title		
3	Article Copyright - Year	**The Author(s) 2020 (This will be the copyright line in the final PDF)**	
4	Journal Name	Current Robotics Reports	
5	Corresponding Author	Family Name	**Amoroso**
6		Particle	
7		Given Name	**Daniele**
8		Suffix	
9		Organization	University of Cagliari
10		Division	Department of Law
11		Address	Cagliari, Italy
12		e-mail	daniele.amoroso@unica.it
13	Author	Family Name	**Tamburrini**
14		Particle	
15		Given Name	**Guglielmo**
16		Suffix	
17		Organization	University of Naples Federico II
18		Division	Department of Electrical Engineering and Information Technology
19		Address	Naples, Italy
20		e-mail	tamburrini@unina.it
21	Schedule	Received	
22		Revised	
23		Accepted	
24	Abstract	**Purpose of Review:** To provide readers with a compact account of ongoing academic and diplomatic debates about autonomy in weapons systems, that is, about the moral and legal acceptability of letting a robotic system to unleash destructive force in warfare and take attendant life-or-death decisions without any human intervention. **Recent Findings:** A précis of current debates is provided, which focuses on the requirement that all weapons systems, including autonomous ones, should remain under meaningful human control (MHC) in order to be ethically acceptable and lawfully employed. Main approaches to MHC are described and briefly analyzed, distinguishing between uniform, differentiated, and prudential	

AUTHOR'S PROOF

policies for human control on weapons systems.
Summary: The review highlights the crucial role played by the robotics research community to start ethical and legal debates about autonomy in weapons systems. A concise overview is provided of the main concerns emerging in those early debates: respect of the laws of war, responsibility ascription issues, violation of the human dignity of potential victims of autonomous weapons systems, and increased risks for global stability. It is pointed out that these various concerns have been jointly taken to support the idea that all weapons systems, including autonomous ones, should remain under meaningful human control (MHC). Main approaches to MHC are described and briefly analyzed. Finally, it is emphasized that the MHC idea looms large on shared control policies to adopt in other ethically and legally sensitive application domains for robotics and artificial intelligence.

25	Keywords separated by ' - '	Autonomous weapons systems - Roboethics - International humanitarian law - Human-robot shared control - Meaningful human control
26	Foot note information	This article is part of the Topical Collection on *Roboethics* Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Current Robotics Reports
https://doi.org/10.1007/s43154-020-00024-3

ROBOETHICS (G VERUGGIO, SECTION EDITOR)

Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues

Daniele Amoroso[1] · **Guglielmo Tamburrini**[2]

Abstract

Purpose of Review To provide readers with a compact account of ongoing academic and diplomatic debates about autonomy in weapons systems, that is, about the moral and legal acceptability of letting a robotic system to unleash destructive force in warfare and take attendant life-or-death decisions without any human intervention.

Recent Findings A précis of current debates is provided, which focuses on the requirement that all weapons systems, including autonomous ones, should remain under meaningful human control (MHC) in order to be ethically acceptable and lawfully employed. Main approaches to MHC are described and briefly analyzed, distinguishing between uniform, differentiated, and prudential policies for human control on weapons systems.

Summary The review highlights the crucial role played by the robotics research community to start ethical and legal debates about autonomy in weapons systems. A concise overview is provided of the main concerns emerging in those early debates: respect of the laws of war, responsibility ascription issues, violation of the human dignity of potential victims of autonomous weapons systems, and increased risks for global stability. It is pointed out that these various concerns have been jointly taken to support the idea that all weapons systems, including autonomous ones, should remain under meaningful human control (MHC). Main approaches to MHC are described and briefly analyzed. Finally, it is emphasized that the MHC idea looms large on shared control policies to adopt in other ethically and legally sensitive application domains for robotics and artificial intelligence.

Keywords Autonomous weapons systems · Roboethics · International humanitarian law · Human-robot shared control · Meaningful human control

Introduction

Robotics has extensively contributed to modify defense systems. Significant examples from the recent past include teleoperated robots detecting and defusing explosive devices (e.g., PackBot) [1], in addition to unmanned vehicles for reconnaissance and combat missions, operating on the ground (e.g., Guardium [2] or TALON [3]) or in the air (e.g., MQ-1 Predator [4]). The deployment of these military robots has been seldom objected to on ethical or legal grounds, with the notable exception of extraterritorial targeted killings accomplished by means of unmanned aerial vehicles. These targeted killings have raised concerns about the infringement of other States' sovereignty and overly permissive application of lethal force in counter-terrorism operations [5–7].

One should carefully note that the release of destructive force by any weaponized robot in the above list is firmly in the hands of human operators. Accordingly, ethical and legal controversies about these systems were confined to a handful of their specific uses, and their overall acceptability as weapons systems was never questioned. However, the entrance on the scene of autonomous weapons systems (AWS from now on) has profoundly altered this ethical and legal landscape.

To count as autonomous, a weapons system must be able to select and engage targets without any human intervention after its activation [8••, 9, 10]. Starting from this basic and quite inclusive condition, the Stockholm International Peace Research

This article is part of the Topical Collection on *Roboethics*

✉ Daniele Amoroso
daniele.amoroso@unica.it

Guglielmo Tamburrini
tamburrini@unina.it

Q1 [1] Department of Law, University of Cagliari, Cagliari, Italy

[2] Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy

Institute (SIPRI) [11] introduced additional distinctions between types of existing AWS: (i) *air defense systems* (e.g., Phalanx [12], MANTIS [13], Iron Dome [14], Goalkeeper [15]); (ii) *active protection systems*, which shield armored vehicles by identifying and intercepting anti-tank missiles and rockets (e.g., LEDS-150 [16] and Trophy [17]); (iii) *robotic sentries*, like the Super aEgis II stationary robotic platform tasked with the surveillance of the demilitarized zone between North and South Korea [18]; (iv) *guided munitions*, which autonomously identify and engage targets that are not in sight of the attacking aircraft (e.g., the Dual-Mode Brimstone [19]); and (v) *loitering munitions*, such as the Harpy NG [20], which overfly an assigned area in search of targets to dive-bomb and destroy.

This classification stands in need of continual expansion on account of ongoing military research projects on unmanned ground, aerial, and marine vehicles that are capable of autonomously performing targeting decisions. Notably, research work based on swarm intelligence technologies is paving the way to swarms of small-size and low-cost unmanned weapons systems. These are expected to overwhelm enemy defenses by their numbers and may additionally perform autonomously targeting functions [21–24].

The technological realities and prospects of AWS raise a major ethical and legal issue: Is it permissible to let a robotic system unleash destructive force and take attendant life-or-death decisions without any human intervention? This issue prompted intense and ongoing debates, at both academic and diplomatic levels, on the legality of AWS under international law [25]. An idea that has rapidly gained ground across the opinion spectrum in this debate is that all weapons systems, including autonomous ones, should remain under meaningful human control (MHC) in order to be ethically acceptable and lawfully employed (see the reports by the UK-based NGO Article 36 [26, 27]). Nevertheless, the precise normative content of such requirement is still far from being precisely spelled out and agreed upon.

This review provides a general survey of the AWS debate, focusing on the MHC turning point and its ethical and legal underpinnings. After recalling the initial stages of the debate, a schematic account is provided of chief ethical and legal concerns about autonomy in weapons systems. Then, the main proposals regarding the MHC content are introduced and analyzed, including our own proposal of a "differentiated and prudential" human control policy on AWS. Finally, it is pointed out how our proposal may help overcome the hurdles that are currently preventing the international community from adopting a legal regulation on the matter.

Highlights from the AWS Ethical and Legal Debate

Members of the robotics community, notably Ronald C. Arkin and Noel Sharkey, were chief protagonists of early discussions about the ethical and legal acceptability of AWS. Arkin emphasized some ethical pros of autonomy in weapons systems. He was concerned about the poor record of human compliance with international norms governing the conduct of belligerent parties in warfare (Laws of War or international humanitarian law (IHL)). In his view, this state of affairs ultimately depends on human self-preservation needs and emotional reactions in the battlefield—fear, anger, frustration, and so on—that a robot is immune to. Arkin's own research on military applications of robotics was inspired by a vision of "ethically restrained" autonomous weapons systems that are capable of abiding "by the internationally agreed upon Laws of War" better than human warfighters. He presented this vision and its ethical motivations in an invited talk at the First International Symposium on Roboethics, organized by Scuola di Robotica, chaired by Gianmarco Veruggio, and held in 2004 at Villa Alfred Nobel in Sanremo, Italy. Arkin later described this meeting as "a watershed event in robot ethics" [28••, 29, 30].

In contrast with Arkin's views, Sharkey emphasized various ethical cons of autonomy in weapons systems. He argued that foreseeable technological developments of robotics and artificial intelligence (AI) offer no support for the idea of autonomous robots ensuring a better-than-human application of the IHL principles. He emphasized that interactions among AWS in unstructured warfare scenarios would be hardly predictable and fast enough to bring the pace of war beyond human control. And he additionally warned that AWS threaten peace at both regional and global levels by making wars easier to wage [31–34]. Sharkey co-founded the International Committee for Robot Arms Control (ICRAC) in 2009 and played a central role in creating the conditions for launching the Campaign to Stop Killer Robots. This initiative is driven by an international coalition of non-governmental organizations (NGOs), formed in 2012 with the goal of "preemptively ban[ning] lethal robot weapons that would be able to select and attack targets without any human intervention."

A similar call against "offensive autonomous weapons beyond meaningful human control" was made in the "Open Letter from AI & Robotic Researchers," released in 2015 by the Future of Life Institute and signed by about 4500 AI/robotics researchers and more than 26,000 other persons, including many prominent scientists and entrepreneurs. Quite remarkably, the Open Letter urges AI and robotics researchers to follow in the footsteps of those scientists working in biology and chemistry, who actively contributed to the initiatives that eventually led to international treaties prohibiting biological and chemical weapons [35].

Worldwide pressures from civil society prompted States to initiate discussion of normative frameworks to govern the design, development, deployment, and use of AWS. Diplomatic dialogs on this topic have been conducted since 2014 at the United Nations in Geneva, within the institutional framework of the Convention on Certain Conventional Weapons (CCW). The CCW's main purpose is to restrict and possibly ban the use of weapons that are deemed to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately. Informal Meetings of Experts on lethal autonomous weapons systems were held on an annual basis at the CCW in Geneva, from 2014 to 2016. Subsequently, the CCW created a Group of Governmental Experts (GGE) on lethal autonomous weapons systems (LAWS), which still remains (as of 2020) the main institutional forum where the issue of autonomy in weapons systems is annually debated at an international level. Various members of the robotics research community take part to the GGE's meetings. So far, the main outcome of the GGE's work is the adoption by consensus of a non-binding instrument, that is, the 11 Guiding Principles on LAWS, which include broad recommendations on human responsibility (Principles (b) and (d)) and human-machine interaction (Principle (c)) [36].

A clear outline of the main ethical and legal concerns raised by AWS is found already in a 2013 report, significantly devoted to "lethal autonomous robotics and the protection of life," by the UN Special Rapporteur on extrajudicial, summary, or arbitrary executions, Christof Heyns [37••]. These concerns are profitably grouped under four headings: (i) *compliance with IHL*, (ii) *responsibility ascription problems*, (iii) *violations of human dignity*, and (iv) *increased risk for peace and international stability*. Let us briefly expand on each one of them, by reference to relevant sections in Heyns' report.

(i) Compliance with IHL would require capabilities that are presently possessed by humans only and that no robot is likely to possess in the near future, i.e., to achieve situational awareness in unstructured warfare scenarios and to formulate appropriate judgments there (paras. 63–74) (in the literature, see [38–40] for a critique of this argument and [41–43] for a convincing rejoinder).
(ii) Autonomy in weapons systems would hinder responsibility ascriptions in case of wrongdoings, by removing human operators from the decision-making process (paras. 75–81) (for further discussion, see [44–46]).
(iii) The deployment of lethal AWS would be an affront to human dignity, which dictates that decisions entailing human life deprivation should be reserved to humans (paras. 89–97) (see [47–49] for more in-depth analysis, as well as [50] for a critical perspective).
(iv) Autonomy in weapons systems would threaten in special ways international peace and stability, by making wars easier to wage on account of reduced numbers of involved soldiers, by laying the conditions for unpredictable interactions between AWS and their harmful outcomes, and by accelerating the pace of war beyond human reactive abilities (paras. 57–62) (this point has been further elaborated in [51]).

These sources of concern jointly make the case for claiming that a meaningful human control (MHC) over weapons systems should be retained exactly in the way of their critical target selection and engagement functions. Accordingly, the notion of MHC enters the debate on AWS as an ethically and legally motivated constraint on the use of any weapons systems, including autonomous ones. The issue of human-robot shared control in warfare is thereby addressed from a distinctive humanitarian perspective, insofar as autonomous targeting may impinge, and deeply so, upon the interests of persons and groups of persons that are worthy of protection from ethical or legal standpoints.

But what does MHC more precisely entail? What is normatively demanded to make human control over weapons systems truly "meaningful"? The current debate about AWS, which we now turn to consider, is chiefly aimed to provide an answer to these questions.

Uniform Policies for Meaningful Human Control

The foregoing ethical and legal reasons go a long way towards shaping the content of MHC, by pinpointing general functions that should be prescriptively assigned to humans in shared control regimes and by providing general criteria to distinguish perfunctory from truly meaningful human control. More specifically, the ethical and legal reasons for MHC suggest a threefold role for human control on weapons systems to be "meaningful." First, the obligation to comply with IHL entails that human control must play the role of a *fail-safe actor*, contributing to prevent a malfunctioning of the weapon from resulting in a direct attack against the civilian population or in excessive collateral damages [52••]. Second, in order to avoid accountability gaps, human control is required to function as *accountability attractor*, i.e., to secure the legal conditions for responsibility ascription in case a weapon follows a course of action that is in breach of international law. Third and finally, from the principle of human dignity respect, it follows that human control should operate as a *moral agency enactor*, by ensuring that decisions affecting the life, physical integrity, and property of people (including combatants) involved in armed conflicts are not taken by non-moral artificial agents [53].

But how are human-weapon partnerships to be more precisely shaped on the basis of these broad constraints? Several attempts to answer this question have been made by parties involved in the AWS ethical and legal debate. The answers that we turn to examine now outline *uniform* human control policies, whereby one size of human control is claimed to fit all AWS and each one of their possible uses. These are the "boxed autonomy," "denied autonomy," and "supervised autonomy" control policies.

The boxed autonomy policy assigns to humans the role of constraining the autonomy of a weapons system within an operational box, constituted by "predefined [target] parameters, a fixed time period and geographical borders" [54]. Accordingly, the weapons system would be enabled to autonomously perform the critical functions of selecting and engaging targets, but only within the boundaries set forth by the human operator or the commander at the planning and activation stages [55–57].

The boxed autonomy policy seems to befit a variety of *deliberate targeting* situations, which involve military objectives that human operators know in advance and can map with high confidence within a defined operational theater. It seems, however, unsuitable to govern a variety of *dynamic targeting* situations. These require one to make changes on the fly to planned objectives and to pursue targets of opportunity. The latter are unknown to exist in advance (unanticipated targets) or else are not localizable in advance with sufficient precision in the operational area (unplanned targets). Under these conditions, boxed autonomy appears to be problematic from a normative perspective, insofar as issues of distinction and proportionality that one cannot foresee at the activation stage may arise during mission execution.

By the same token, a boxed autonomy policy may not even suffice to govern deliberate targeting of military objectives placed in unstructured warfare scenarios. To illustrate, consider the loitering munition Harpy NG, endowed with the capability of patrolling for several hours a predefined box in search of enemy targets satisfying given parameters. The conditions licensing the activation of this loitering munition may become superseded if civilians enter the boxed area, erratic changes occur, or surprise-seeking intentional behaviors are enacted [58]. Under these various circumstances, there is "fail-safe" work for human control to do at the mission execution stage too.

In sharp contrast with the boxed autonomy policy, the denied autonomy policy rules out any autonomy whatsoever for weapons systems in the critical targeting function and therefore embodies a most restrictive interpretation of MHC [59]. Denied autonomy undoubtedly fulfills the threefold normative role for human control as fail-safe actor, accountability attractor, and moral agency enactor. However, this policy has been sensibly criticized for setting too high a threshold for machine autonomy, in ways that are divorced from "the reality of warfare and the weapons that have long been considered acceptable in conducting it" [60]. To illustrate this criticism, consider air defensive systems, which autonomously detect, track, and target incoming projectiles. These systems have been aptly classified as SARMO weapons, where SARMO stands for "Sense and React to Military Objects." SARMO systems are hardly problematic from ethical and legal perspectives, in that "they are programmed to automatically perform a small set of defined actions repeatedly. They are used in highly structured and predictable environments that are relatively uncluttered with a very low risk of civilian harm. They are fixed base, even on Naval vessels, and have constant vigilant human evaluation and monitoring for rapid shutdown" [61].

SARMO systems expose the overly restrictive character of a denied autonomy policy. Thus, one wonders whether milder forms of human control might be equally able to strip the autonomy of weapons systems of its ethically and legally troubling implications. This is indeed the aim of the supervised autonomy policy, which occupies a middle ground between boxed and denied autonomy, insofar as it requires humans to be on the loop of AWS missions.

As defined in the US DoD Directive 3000.09 on "Autonomy in Weapons Systems," human-supervised AWS are designed "to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur" (p. 13). Notably, human-supervised AWS may be used for defending manned installations and platforms from "attempted time-critical or saturation attacks," provided that they do not select "humans as targets" (p. 3, para. 4(c)(2); see, e.g., the Phalanx Close-In Weapons System in use on the US surface combat ships). While undoubtedly effective for these and other warfare scenarios, supervised autonomy is not the silver bullet for every ethical and legal concern raised by AWS. To begin with, by keeping humans on-the-loop, one would not prevent faster and faster offensive AWS from being developed, eventually reducing the role of human operators to a perfunctory supervision of decisions taken at superhuman speed while leaving the illusion that the human control requirement is still complied with [62]. Moreover, the automation bias—the human propensity to overtrust machine decision-making processes and outcomes—is demonstrably exacerbated by a distribution of control privileges that entrusts humans solely with the power of overriding decisions autonomously taken by the machines [63].

To sum up, each one of the boxed, denied, and supervised autonomy policies provides useful hints towards a normatively adequate human-machine shared control policy for military target selection and engagement. However, the complementary defects of these *uniform* control policies suggest the implausibility of solving the MHC problem with one formula, to be applied to all kinds of weapons systems and to each one of their possible uses. This point was consistently made by the US delegation at GGE meetings in Geneva: "there is not a fixed, one-size-fits-all level of human judgment that should be applied to every context" [64].

Differentiated Policies for Meaningful Human Control

Other approaches to MHC aim to reconcile the need for differentiated policies with the above ethical and legal constraints on human control. Differentiated policies modulate human control along various autonomy levels for weapons systems. Autonomy levels have been introduced in connection with, say, automated driving, surgical robots, and unmanned commercial ships to discuss technological roadmaps or ethical and legal issues [65–67]. A taxonomy of increasing autonomy levels concerning the AWS critical target selection and engagement functions was proposed by Noel Sharkey (and only slightly modified here, with regard to levels 4 and 5) [68••].

L1. A human engages with and selects targets and initiates any attack.
L2. A program suggests alternative targets, and a human chooses which to attack.
L3. A program selects targets, and a human must approve before the attack.
L4. A program selects and engages targets but is supervised by a human who retains the power to override its choices and abort the attack.
L5: A program selects targets and initiates attack on the basis of the mission goals as defined at the planning/activation stage, without further human involvement.

The main uniform control policies, including those examined in the previous section, are readily mapped onto one of these levels.

L5 basically corresponds to the boxed autonomy policy, whereby MHC is exerted by human commanders at the planning stage of the targeting process only. As noted above, boxed autonomy does not constitute a sufficiently comprehensive and normatively acceptable form of human-machine shared control policy.

L4 basically corresponds to the supervised autonomy policy. The uniform adoption of this level of human control must also be advised against in the light of automation bias risks and increasing marginalization of human oversight. In certain operational conditions, however, it may constitute a normatively acceptable level of human control.

L3 has been seldom discussed in the MHC debate. At this level, control privileges on critical targeting functions are equally distributed between weapons system (target selection) and human operator (target engagement). To the extent that the human deliberative role is limited to approving or rejecting targeting decisions suggested by the machine, this level of human control does not provide adequate bulwarks against the risk of automation bias [69]. In the same way as L4, therefore, it should not be adopted as a general policy.

L1 and L2 correspond to shared control policies where the weapons system's autonomy is either totally absent (L1) or limited to the role of adviser and decision support system for human deliberation (L2). The adoption of these pervasive forms of human control must also be advised against insofar as some weapons (notably SARMO systems) have long been considered acceptable in warfare operations.

In the light of these difficulties, one might be tempted to conclude that the search for a comprehensive and normatively binding MHC policy should be given up and that the best one can hope for is the exchange of good practices between States about AWS control, in addition to the proper application of national mechanisms to review the legality of weapons [70–72]. But alternatives are possible, which salvage the idea of a comprehensive MHC policy, without neglecting the need for differentiated levels of AWS autonomy in special cases. Indeed, the authors of this review have advanced the proposal of a comprehensive MHC policy, which is jointly differentiated and *prudential* [73, 74].

The prudential character of this policy is embodied into the following default rule: low levels of autonomy L1–L2 should be exerted on all weapons systems and uses thereof, unless the latter are included in a list of exceptions agreed on by the international community of States. The prudential imposition by default of L1 and L2 is aimed at minimizing the risk of breaches of IHL, accountability gaps, or affronts to human dignity, should international consensus be lacking on whether, in relation to certain classes of weapons systems or uses thereof, higher levels of machine autonomy are equally able to grant the fulfillment of genuinely meaningful human control. The differentiated character of this policy is embodied in the possibility of introducing internationally agreed exceptions to the default rule. However, these exceptions should come with the indication of what level is required to ensure that the threefold role of MHC (fail-safe actor, accountability attractor, moral agency enactor) is adequately performed.

In the light of the above analysis, this should be done by taking into account at least the following observations:

1. The L4 human supervision and veto level might be deemed as an acceptable level of control only in case of anti-materiel AWS with exclusively defensive functions (e.g., Phalanx or Iron Dome). In this case, ensuring that human operators have full control over every single targeting decision would pose a serious security risk, which makes the application of L1, L2, and L3 problematic from both military and humanitarian perspectives. The same applies to active protection systems, like Trophy, provided that their use in supervised-autonomy mode is excluded in operational environments involving a high concentration of civilians.
2. L1 and L2 could also be impracticable in relation to certain missions because communication constraints would allow only limited bandwidth. In this case, military considerations should be balanced against humanitarian ones. One might allow for less bandwidth-heavy

(L3) control in two cases: deliberate targeting and dynamic targeting in fully structured scenarios, e.g., in high seas. In both hypotheses, indeed, the core targeting decisions have actually been taken by humans at the planning/activation stage. Unlike L4, however, L3 ensures that there is a human on the attacking end who can verify, in order to deny or grant approval, whether there have been changes in the battlespace which may affect the lawfulness of the operation. Looking at existing technologies, L3 might be applied to sentry robots deployed in a fully structured environment, like the South Korean Super aEgis II.

3. The L5 boxed autonomy level should be considered incompatible with the MHC requirement, unless operational space and time frames are so strictly circumscribed to make targeting decisions entirely and reliably traceable to human operators.

Concluding Remarks

Recent advances in autonomous military robotics have raised unprecedented ethical and legal issues. Regrettably, diplomatic discussions at the GGE in Geneva not only have so far fallen short of working out a veritable legal regime on meaningful human control over AWS, but—what is worse—are currently facing a stalemate, which is mainly determined by the opposition of major military powers, including the US and the Russian Federation, to the adoption of any kind of international regulation on the matter.

Our proposal of relinquishing the quest for a one-size-fits-all solution to the MHC issue in favor of a suitably differentiated approach may help sidestep current stumbling blocks. Diplomatic and political discontent about an MHC requirement that is overly restrictive with respect to the limited autonomy of some weapons systems might indeed be mitigated recognizing the possibility of negotiating exceptions to L1–L2 human control, by identifying weapons systems and contexts of use where milder forms of human control will suffice to ensure the fulfillment of the fail-safe, accountability, and moral agency properties whose preservation generally underpins the normative concerns about weapons' autonomy in targeting critical functions.

In a broader perspective, a differentiated approach to MHC may be of some avail as regards the general issue of human control over intelligent machines operating in ethically and legally sensitive domains, insofar as the MHC language has been recently used about autonomous vehicles [75, 76] and surgical robots [77].

Funding Information Open access funding provided by Università degli Studi di Cagliari within the CRUI-CARE Agreement.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Rudakevych P, Ciholas M. PackBot EOD firing system. SPIE Proceedings, Volume 5804, Unmanned Ground Vehicle Technology VII. 2005.
2. Egozi A. Robotics: Israel Aerospace Industries (IAI) ground robotic systems. Asia-Pac Defence Rep. 2016;42(7):46–7.
3. Wells P, Deguire D. TALON: a universal unmanned ground vehicle platform, enabling the mission to be the focus. SPIE Proceedings, Volume 5804, Unmanned Ground Vehicle Technology VII. 2005.
4. Visnevski NA, Castillo-Effen M. A UAS capability description framework: reactive, adaptive, and cognitive capabilities in robotics. 2009 IEEE Aerospace Conference, Big Sky, Montana. 2009.
5. Alston P. Report of the United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions. Addendum. Study on targeted killings. UN Doc. A/HRC/14/24/Add.6. 28 May 2010.
6. O'Connell ME. The choice of law against terrorism. J Natl Sec Law Policy. 2009;4:343–68.
7. Melzer N. Human rights implications of the usage of drones and unmanned robots in warfare. Report requested by the European Parliament's Subcommittee on Human Rights. EXPO/B/DROI/2012/12. May 2013.
8.•• US Department of Defense. Directive 3000.09 "Autonomy in Weapons Systems". 21 November 2012. **The Directive crucially contributed to the AWS debate by introducing a workable definition of AWS and by establishing the principle that "appropriate levels of human judgment over the use of force" should always be ensured over weapons systems**.
9. International Committee of the Red Cross (ICRC). Views on autonomous weapon system. Paper submitted to the Informal meeting of experts on lethal autonomous weapons systems of the

Convention on Certain Conventional Weapons (CCW), Geneva. 11 April 2016.

10. Campaign to Stop Killer Robots. Urgent Action Needed to Ban Fully Autonomous Weapons. Press release. 23 April 2013.
11. Boulanin V, Verbruggen N. Mapping the development of autonomy in weapon systems. Solna: SIPRI Report; 2017.
12. Stoner RH. R2D2 with attitude: the story of the phalanx close-in weapons. 2009. Available at: www.navweaps.com/index_tech/tech-103.htm.
13. NBS MANTIS Air Defence Protection System. Available at: https://www.army-technology.com/projects/mantis/.
14. Landau EB, Bermant A. Iron dome protection: missile defense in Israel's security concept. In: Kurz A, Brom S, editors. The lessons of operation protective edge. Tel Aviv: Institute for National Security Studies; 2014. p. 37–42.
15. 30 mm (1.2″) Goalkeeper SGE-30. Available at: http://www.navweaps.com/Weapons/WNNeth_30mm_Goalkeeper.php.
16. SAAB Group. LEDS full spectrum active protection for land vehicles Available at: https://saab.com/globalassets/commercial/land/force-protection/active-protection/leds/leds-product-sheet.pdf.
17. Trophy Active Protection System; 10 April 2007. Available at: https://defense-update.com/20070410_trophy-2.html.
18. Parkin S. Killer Robots: the soldiers that never sleep. BBC Future; 16 July 2015. Available at: www.bbc.com/future/story/20150715-killer-robots-the-soldiers-that-never-sleep.
19. UK Royal Air Force. Aircraft & weapons. 2007: p. 87.
20. Gettinger D, Michel AH. Loitering munitions. Center for the Study of the Drone; 2017.
21. Scharre P. Robotics on the battlefield part II. The coming swarm. Center for a New American Security; October 2014.
22. Brehm M, De Courcy Wheele A. Swarms. Article36 discussion paper for the Convention on Certain Conventional Weapons (CCW), Geneva. March 2019.
23. Verbruggen M. The question of swarms control: challenges to ensuring human control over military swarms. Non-Proliferation and Disarmament Paper No. 65; December 2019.
24. Ekelhof MAC, Persi Paoli G. Swarm robotics: technical and operational overview of the next generation of autonomous systems. United Nations Institute for Disarmament Research (UNIDIR); 2020.
25. Amoroso D. Autonomous Weapons Systems and International Law. A study on human-machine interactions in ethically and legally sensitive domains. Naples/Munich: ESI/Nomos; 2020.
26. Roff HM, Moyes R. Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. Article36 briefing paper prepared for the CCW informal meeting of experts on lethal autonomous weapons systems. April 2016.
27. Moyes R. Key elements of meaningful human control. Article 36 Background paper to comments prepared for the CCW Informal Meeting of Experts on Lethal Autonomous Weapons Systems. April 2016.
28.•• Arkin RC. Governing Lethal Behavior in Autonomous Robots. Boca Raton: CRC Press; 2009. **Arkin's book represents the first—and, so far, the best articulated—attempt by a roboticist to argue for the desirability of AWS from an ethical and legal perspective. It deeply influenced the subsequent debate by providing major military powers with a convenient argument to assert the legality of AWS**.
29. Arkin RC. Lethal autonomous systems and the plight of the non-combatant. AISB Quarterly. 2013;137:1–9.
30. Arkin RC. A roboticist's perspective on lethal autonomous weapon systems. In: Perspectives on lethal autonomous weapon systems. UNODA Occasional Papers No 30; November 2017: p. 35–47.
31. Sharkey NE. Automated killers and the computing profession. Computer. 2007;40(11):122–3.
32. Sharkey NE. Cassandra or false prophet of doom: AI robots and war. IEEE Intell Syst. 2008;23(4):14–7.
33. Sharkey NE. Saying 'no!' to lethal autonomous targeting. J Military Ethics. 2010;9(4):369–83.
34.• Sharkey NE. The evitability of autonomous robot warfare. Int Rev Red Cross. 2012;94:787–9 **This article provides the most comprehensive outline, from a roboticist's perspective, of the ethical and legal concerns raised by autonomy in weapons systems**.
35. Autonomous weapons: an open letter from AI & robotics researchers. 28 July 2015. Available at: https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1.
36. Final Report of the 2019 Meeting of the high contracting parties to the CCW. UN Doc CCW/MSP/2019/CRP2/Rev1, Annex III 15 November 2019.
37.•• Heyns C. Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions. UN Doc. A/HRC/23/47. 9 April 2013. **Heyns' report is a milestone in the AWS debate: it raised awareness within the United Nations as to the ethical and legal implications of autonomy in weapons systems, by illustrating—concisely, but also comprehensively—the main normative issues at stake**.
38. Schmitt MN, Thurnher JS. "Out of the loop": autonomous weapon systems and the law of armed conflict. Harvard Natl Secur J. 2013;4:231–81.
39. Sassòli M. Autonomous weapons and international humanitarian law: advantages, open technical questions and legal issues to be clarified. Int Law Stud. 2014;90:308–40.
40. Anderson K, Waxman M. Debating autonomous weapon systems, their ethics, and their regulation under international law. In: Brownsword R, Scotford E, Yeung F, editors. The Oxford handbook of law, regulation, and technology. New York: Oxford University Press; 2017. p. 1097–117.
41. Krupyi T. Of souls, spirits and ghosts: transposing the application of the rules of targeting to lethal autonomous robots. Melbourne J Int Law. 2015;16(1):145–202.
42. Brehm M. Defending the boundary: constraints and requirements on the use of autonomous weapon systems under international humanitarian and human rights law. Geneva Academy of International Humanitarian Law and Human Rights, Academy briefing No. 9: May 2017.
43. Geiss R, Lahmann H. Autonomous weapons systems: a paradigm shift for the law of armed conflict? In: Ohlin JD, editor. Research handbook on remote warfare. Cheltenham: Edward Elgar Publishing; 2017. p. 371–404.
44.• Sparrow R. Killer robots. J Appl Philos. 2007;24(1):62–77 **Interested readers find here a first, ground-breaking illustration of the risk that increasing autonomy in weapons system creates "accountability gaps" in case of unlawful targeting decisions**.
45. Amoroso D, Giordano B. Who is to blame for autonomous weapons systems' misdoings? In: Lazzerini N, Carpanelli E, editors. Use and misuse of new technologies. Contemporary Challenges in International and European Law. The Hague: Springer; 2019. p. 211–32.
46. McDougall C. Autonomous weapon systems and accountability: putting the cart before the horse. Melbourne J Int Law. 2019;20(1):58–87.
47. Asaro P. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. Int Rev Red Cross. 2012;94:687–709.
48. Leveringhaus A. Ethics and autonomous weapons. London: Palgrave; 2016.
49. Heyns C. Autonomous weapons in armed conflict and the right to a dignified life: an African perspective. South African J Hum Rights. 2017;33(1):46–71.

50. Birnbacher D. Are Autonomous weapons systems a threat to human dignity? In: Bhuta N, et al., editors. Autonomous weapons systems. Law, ethics, policy. Cambridge: Cambridge University Press; 2016. p. 105–21.
51. Altmann J, Sauer F. Autonomous weapon systems and strategic stability. Survival. 2017;59(5):117–42.
52.•• Scharre P. Army of none. Autonomous weapons and the future of war. New York/London: W.W. Norton; 2018. **This volume represents one of the few book-size studies taking stock of the discussion and attempting to delineate workable solutions to the problems raised by autonomy in weapons systems. As an added value, it has been written by a key-actor in the AWS debate, since Scharre coordinated the Drafting process that eventually brought about the US DoD Directive 3000.09**.
53. ICRC. Ethics and autonomous weapon systems: an ethical basis for human control? Working paper submitted to the Group of Governmental Experts on lethal autonomous weapons of the CCW, Geneva. UN Doc. CCW/GGE.1/2018/WP.5. 29 March 2018.
54. International Panel on the Regulation of Autonomous Weapons (iPRAW). Focus on Human Control. August 2019.
55. Dutch Advisory Council on International Affairs (AIV) and Advisory Committee on Issues of Public International Law (CAVV). Report on Autonomous weapon systems: the need for meaningful human control. No. 97 AIV / No. 26 CAVV. 2015.
56. Roorda M. NATO's targeting process: ensuring human control over and lawful use of 'autonomous' weapons. In: Williams A, Scharre P, editors. Autonomous systems: issues for defence policymakers. The Hague: NATO Communications and Information Agency; 2015. p. 152–68.
57. Ekelhof MAC. Moving beyond semantics on autonomous weapons systems: meaningful human control in operation. Glob Policy. 2019;10(3):343–8.
58. Akerson D. The illegality of offensive lethal autonomy. In: Saxon D, editor. International humanitarian law and the changing technology of war. Leiden/Boston: Brill/Nijhoff; 2013. p. 65–98.
59. Chengeta T. Defining the emerging notion of 'meaningful human control' in autonomous weapon systems. New York J Int Law Politics. 2017;49:833–90.
60. Horowitz MC, Scharre P. Meaningful human control in weapon systems: a primer, CNAS Working Paper, March 2015.
61. International Committee for Robot Arms Control. Statement on technical issues. Informal meeting of experts on lethal autonomous weapons, Geneva. 14 May 2014.
62. Schwarz E. The (Im)possibility of meaningful human control for lethal autonomous weapon systems. Humanitarian Law & Policy (Blog of the ICRC). 29 August 2018.
63. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? Int J Hum-Comput Stud. 1999;51(5):991–1006.
64. United States. Human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems. Working paper submitted to the Group of Governmental Experts on lethal autonomous weapons of the CCW, Geneva. UN Doc. CCW/GGE.2/2018/WP.4. 28 August 2018.
65. SAE International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. 15 June 2018. Available at: https://www.sae.org/standards/content/j3016_201806/.
66. Yang G-Z et al. Medical robotics – regulatory, ethical, and legal considerations for increasing levels of autonomy. Sci Robot. 2017:2(4).
67. DNV GL. Autonomous and remotely operated ships. September 2018.
68.•• Sharkey NE. Staying the Loop: human supervisory control of weapons. In: Bhuta N, et al., editors. Autonomous weapons systems. Law, ethics, policy. Cambridge: Cambridge University Press; 2016. p. 23–38. **Sharkey's chapter offers a valuable comparison of strengths and weaknesses of human and machine decision-making processes, paying particular attention to implications for autonomous robotics in military applications. Also, by providing a clear taxonomy of possible forms of human-weapon partnership, it laid the groundwork for subsequent research on MHC**.
69. Cummings ML. Automation and accountability in decision support system interface design. J Technol Stud. 2006:23–31.
70. Switzerland. A "compliance-based" approach to autonomous weapon systems. Working paper submitted to the Group of Governmental Experts on lethal autonomous weapons of the CCW, Geneva. UN Doc. CCW/GGE.1/2017/WP.9. 10 November 2017.
71. United States. Statement for the General Exchange of Views. Group of Governmental Experts on lethal autonomous weapons of the CCW, Geneva. 9 April 2018.
72. Israel. Statement on Agenda Item 6(d). Group of Governmental Experts on lethal autonomous weapons of the CCW, Geneva. 29 August 2018.
73. Amoroso D, Tamburrini G. What makes human control over weapon systems "meaningful"? ICRAC working paper #4. August 2019.
74. Amoroso D, Tamburrini G. Filling the empty box: a principled approach to meaningful human control over weapons systems. ESIL Reflections. 2019:8(5).
75. Santoni de Sio F, Van Den Hoven J. Meaningful human control over autonomous systems: a philosophical account. Front Robotic AI. 2018.
76. Mecacci G, Santoni de Sio F. Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. Ethics Inf Technol. 2020;22:103–15.
77. Ficuciello F, et al. Autonomy in surgical robots and its meaningful human control. Paladyn J Behav Robot. 2019;10(1):30–43.

AUTHOR QUERY

AUTHOR PLEASE ANSWER QUERY.

Q1. Please check if the provided addresses in the affiliations are correct. Otherwise, please provide the correct address of affiliations 1 and 2.