

Machine autonomy and the constant care obligation

December 11, 2018, Autonomous Weapons / Conduct of Hostilities / Law and Conflict / New Technologies / Weapons

Chris Jenks & Rain Liivoja



The debate about the way the international community should deal with autonomous weapon systems has been ongoing for some five years now. During this time, the focus has been on autonomy in what the ICRC has referred to as the ‘critical functions’ of a weapon system—its capacity to select and attack targets. We argue here that while a weapon system focus is appropriate for the discussions occurring in the context of the Convention on Certain Conventional Weapons (CCW), the application of international humanitarian law (IHL) to autonomy in weapon systems is much broader, extending also to ‘non-critical functions’.

Questions on autonomous weapons were first formally raised in the context of the CCW during the 2013 *Meeting of the States Parties*. Since then, the States Parties—along with civil society and with input from numerous experts—have discussed autonomous weapons at least annually. The latest round, in the form of a Group of Governmental Experts (GGE) meeting, took place in Geneva from 27 until 31 August. In the early morning hours of 1 September, the GGE *affirmed* by consensus ten ‘possible guiding principles’ governing the autonomous weapons discussions and recommended that the discussions continue in 2019. Late in the evening of 23 November, the High Contracting Parties (HCP) to the CCW adopted a final report which included 2019 discussions.

The first of those principles is that ‘[IHL] continues to apply fully to all weapons systems, including the potential development and use of [autonomous weapons]’. There is no disputing that principle, although there was quibbling within the GGE as to how it should be formulated. What remains very much debated, however, is whether existing IHL is adequate for addressing the concerns raised by autonomous weapons, or whether further regulation is necessary.

Scope of Autonomy

The focus of the debate has been on the ‘sharp end’ of autonomy. States Parties and other participants have broadly concerned themselves with weapon systems that have the ability to ‘select and attack’ (e.g., ICRC, UNIDIR) or ‘select and engage’ (e.g., US, UK, Norway, Pakistan, Estonia, HRW) targets without human intervention. But obviously weapon systems can, and do, have non-critical functions with some degree of autonomy. Similarly, non-weaponized military platforms can, and do, perform functions with some degree of autonomy. For example, an unmanned aerial vehicle may have autonomous take-off and landing, navigation, refueling, intelligence, surveillance and reconnaissance and other functions. If that vehicle is weaponized, but the weapon is remote-controlled by a human operator, autonomy would not be in the critical functions.

To be sure, there is some uncertainty as to the meaning of the word ‘autonomous’ in this context. With respect to the critical functions, the lack of a precise meaning of autonomy has been the source of frustration and confusion at the various CCW autonomous weapons meetings. As one of us *has previously written*, ‘the international community cannot even agree about what they disagree about’. Attempts to redefine the problem through the notion of human control have not been successful either. As the other one of us *has written*, ‘the furious agreement about the need to maintain human involvement hides a deep disagreement about what that means.’

We look at autonomy quite broadly. At different points in the operations of modern technological systems, weapons and otherwise, some tasks and functions are performed by humans directly, whereas some are performed by the system, a machine or computer that has been programmed and activated by a human. By machine autonomy we simply mean the capacity of a machine to carry out tasks or functions without the need for real-time input or direction from a human operator. This broad understanding of machine autonomy includes at least some systems which in terms of machine complexity might be considered automated or automatic. We adopt this understanding of machine autonomy out of recognition of the dynamic nature of functions within a system. We also believe that because autonomy is better thought of across several different spectrums, attempts at overall system categorization based on only one of the spectrums—machine complexity—lack practical utility.

Coming back to the law, would IHL be relevant to the non-critical functions of weapon systems or other military systems? The answer is ‘yes’.

Scope of the constant care obligation under IHL

Of particular significance here is Article 57(1) of Additional Protocol I, which reads, in its entirety, as follows:

In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.

The constant care obligation is a *general principle*. But, as the International Law Association Study Group on the Conduct of Hostilities in the 21st Century (Study Group)—which met between 2012–2016—aptly reflected in its *final report*, the ‘generality [of the constant care obligation] need not dilute its significance’. Moreover, the ‘use of the word “shall” ... indicates that whatever it is that this provision entails is binding’ on parties to API.

That this obligation forms a sort of preamble to Article 57 can lead to the *perception that it is merely aspirational*, particularly in the light of its very general wording. Article 57(1) has thus been misunderstood as needing to be read—in order to carry legal weight—in conjunction with the remaining paragraphs of this Article, which deal with precautions in attack. This results in an interpretation of the constant care obligation as only applying to attacks. In terms of machine autonomy, this flawed interpretation would mean that Article 57(1) would only apply to the critical functions mentioned above.

We take the contrary view. Namely, we consider that the obligation of constant care in military operations in Article 57(1) has legal effect independent of—and broader in scope than—the rest of Article 57, which only applies to attacks. As the ICRC Commentary to Article 57 *explains*, the words ‘military operations’ in paragraph (1) should be understood to mean ‘any movements, manoeuvres and other activities whatsoever carried out by the armed forces with a view to combat.’ Unlike the rest of the Article, it has been said that paragraph (1) applies to ‘*all domains of warfare and all levels of operations*’. As *one commentator* noted, ‘this broader field of application implies that [Art. 57(1)] can, on its own, give rise to concrete legal obligations’.

Article 57(1) has gained prominence of late in discussions concerning hostile cyber operations. In this context it has been broadly agreed that cyber operations having effects comparable to the use of kinetic weapons constitute attacks, and are thus subject to Article 57(2)–(5)’s provisions on attacks.^[1] There is some uncertainty as to whether IHL would govern cyber operations falling below this threshold. We subscribe to the view that any hostile cyber operations undertaken in the context of an armed conflict would be captured by the ‘constant care’ obligation of Article 57(1).

While constant care is not defined in IHL, the Study Group contended that the ‘[o]bligation to take constant care is best understood as an obligation of conduct, i.e., a positive and continuous obligation aimed at risk mitigation and harm prevention and the fulfillment of which requires the exercise of due diligence’.

Implications of applying the constant care principle to machine autonomy

What then of machine autonomy? The UK Army, for example, has participated in a Coalition Assured Autonomous Resupply demonstration of driverless resupply vehicles.^[2] The goal of such military autonomous vehicles is to ‘one day reduce the burden on and risk to the military user, while improving logistics efficiencies and interoperability’. But as armed conflicts unfortunately increasingly occur in and around civilian populated areas, Article 57(1) would have some application to such vehicles.

In the first instance, Article 57(1) would require that autonomous vehicles be designed and relied upon with the safety of the civilian population in mind. Thus, an autonomous ground vehicle should avoid, for example, injuring civilians or damaging civilian building and infrastructure. Likewise an autonomous aerial vehicle should be capable of avoiding civilian air traffic and not crash into and damage civilian objects upon a failure of the communication link to its operator.

But there are further implications to Article 57(1). What if autonomous vehicles achieved such a degree of sophistication and safety that they were less likely to crash into civilians and civilian property than remotely-piloted vehicles? Arguably, there would be a point where the constant care obligation to spare the civilian population, civilians and civilian objects would *require* States that have such vehicles to use them. The question of

whether States have a duty to develop, acquire or utilize new technologies in order to comply with IHL obligations is a broader question and one which, while having been addressed by *others*, will no doubt receive additional attention as technology progresses.

Implications for the CCW Autonomous Weapons discussion

Does all this mean that the constant care obligation and non-critical functions, such as navigation of autonomous vehicles, should be included in the CCW autonomous weapons discussion? No, it does not. The purpose of the CCW is to ban or restrict the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately. So, it would be hard to see how a non-weapons function falls under the CCW purview.

At the same time, should future CCW regulation of autonomous weapons be divorced from a discussion of the constant care obligation, this may create a situation where States might be prohibited from using some autonomous functions via a CCW Protocol while at the same time—via AP I—could be obligated to employ other autonomous, non-critical, functions. This could result in tension between a future CCW Lethal Autonomous Weapon Systems (LAWS) Protocol *regulating* autonomy and AP I arguably *requiring* autonomy.

Possible incongruity such as this highlights one of implicit challenges of the autonomous weapons GGE. The CCW is a *treaty regime* focused on specific types of weapons. Yet, the LAWS GGE is, in part, trying to determine how potential advances in autonomy—a general technology descriptor—will influence the development of any existing or future weapon system. This is a massive and potentially interminable task. As a result, the recent decision by States Parties to the CCW to continue discussions on autonomous weapons in 2019 was not surprising. On the one hand and as mentioned above, for those discussions to omit non-critical weapons functions would be consistent with CCW's purpose. But on the other hand, failing to consider the implications of the constant care obligation may prove problematic, particularly for CCW High Contracting Parties which are also States Parties to AP I.

The views set out in this article are those of the authors, Chris Jenks and Rain Liivoja, and do not necessarily reflect the position of any government or institution.

Footnotes

[1] See, e.g., *Tallinn Manual 2.0*, especially Rule 93; *US DoD Law of War Manual*, para 16.5.1; *Australia's International Cyber Engagement Strategy*, Annex A.

[2] Systems with various degrees of machine autonomy of course operate in the air and at sea. See, e.g., Lt. Col. Jeremy C. Gottshall & Capt. Richard A. Lozano, *Autonomous Aerial Resupply in the Forward Support Company*, 50 *Army Sustainment* 44 (2018) (describing the use of autonomous aerial systems to resupply military forces) and *Autonomous Warrior 2018*, an Australian Royal Navy initiative testing the integration of air, land, sea and underwater semi-autonomous and unmanned (and unarmed) systems.

Related blog posts

- *The (im)possibility of meaningful human control for lethal autonomous weapon systems* Elke Schwartz
- *The human nature of international humanitarian law* Noel Sharkey
- *human nature of international humanitarian law* Eric Talbot Jensen
- *Autonomous weapons: Operationalizing meaningful human control* Merel Ekelhof
- *Human judgment and lethal decision-making in war* Paul Scharre
- *Autonomous weapon and human control* Tim McFarland
- *Autonomous weapon systems: An ethical basis for human control?* Neil Davison
- *Autonomous weapon systems: A threat to human dignity?* Ariadna Pop
- *Ethics as a source of law: The Martens clause and autonomous weapons* Rob Sparrow
- *Autonomous weapons mini-series: Distance, weapons technology and humanity in armed conflict* Alex Leveringhaus

Key ICRC documents on AWS

- ICRC, *Autonomous weapons: States must agree on what human control means in practice*, November 2018
- VOX News/IRC Video clip on autonomous weapons, May 2018
- ICRC Statement to UN CCW Group of Governmental Experts (GGE), April 2018
- ICRC Statement to UN CCW Group of Governmental Experts (GGE), November 2017
- ICRC Report on Ethics and autonomous weapon systems, April 2018
- Paper on autonomous weapon systems under international humanitarian law, November 2017
- ICRC Expert meeting report, 2016
- ICRC Expert meeting report, 2014

DISCLAIMER: Posts and discussion on the Humanitarian Law & Policy blog may not be interpreted as positioning the ICRC in any way, nor does the blog's content amount to formal policy or doctrine, unless specifically indicated.

Tags: *autonomous weapons, AWS, conduct of hostilities, IHL, LAWs, LOAC, UNCCW*

Share this article    

Comments

There are no comments for now.

Leave a comment

Name *

Email address * This is for content moderation. Your email address will not be made public.

Your comment

Post comment



Get Law & Policy updates

 TWEET

 SHARE