sipri

# ARTICLE 36 REVIEWS

## Dealing with the challenges posed by emerging technologies

VINCENT BOULANIN AND MAAIKE VERBRUGGEN

# ARTICLE 36 REVIEWS

Dealing with the challenges posed by emerging technologies

VINCENT BOULANIN AND MAAIKE VERBRUGGEN

December 2017


STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

**STOCKHOLM INTERNATIONAL**
**PEACE RESEARCH INSTITUTE**

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

The Governing Board is not responsible for the views expressed in the publications of the Institute.




**STOCKHOLM INTERNATIONAL**
**PEACE RESEARCH INSTITUTE**

Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org

# Contents

# Acknowledgements

SIPRI and the authors would like to express sincere gratitude to the Ministry for Foreign Affairs of Sweden, the Ministry of Foreign Affairs of the Netherlands, and the Department for National Defence of Canada for their support of this project.

In addition, the authors wish to thank the participants—panellists and attendees alike—at the SIPRI international conference on Article 36 reviews and emerging technologies. Rather than summarize the various sessions and panels, this report seeks to reflect and synthesize the participants' cross-cutting contributions, in terms of both ideas and analysis. The report was also produced through research that the authors conducted before and after the conference. For more information about the conference, including a comprehensive list of topics and panellists, please see the programme, which is available at the SIPRI website.

Responsibility for the views and information set forth in this report lies entirely with the authors.

# About the authors

Dr Vincent Boulanin (France/Sweden) is a Researcher within Armament and Disarmament at SIPRI and Principal Investigator of the SIPRI project on Article 36 and Emerging Technologies. He works on issues related to the production, use and control of emerging military and security technologies. Before joining SIPRI in 2014, he completed a PhD in Political Science at École des Hautes Études en Sciences Sociales (the School for Advanced Studies in the Social Sciences) in Paris.

Maaike Verbruggen (Netherlands) joined SIPRI in April 2016 to work as a research assistant with the SIPRI research team working on emerging military and security technologies. She left SIPRI in November 2017 to work as a PhD researcher at the Vrije Universiteit in Brussels. Before joining SIPRI, she completed a Master of Philosophy degree in Peace and Conflict Studies at Oslo University.

# Executive summary

The right of states to choose the means and methods of warfare is not unlimited. International law includes both general rules and treaty law that prohibit or restrict specific effects, types of weapons or means and methods of warfare in armed conflicts. As a general rule, international humanitarian law prohibits the use of weapons, means and methods of warfare that cause superfluous injury or unnecessary suffering, or damage military objectives and civilians or civilian objects without distinction. There are also a number of rules under treaty and customary law that ban specific types of weapons (e.g. biological and chemical weapons or blinding laser weapons) or restrict the way in which they are used, such as the 1907 Convention Relative to the Laying of Automatic Submarine Contact Mines. These restrictions and prohibitions are intended to set minimum standards of humanity during armed conflicts.

As a complement to, and reinforcement of, these limitations, international law—specifically Article 36 of the 1977 Additional Protocol to the 1949 Geneva Conventions—imposes a practical obligation on states to determine whether 'in the study, development, acquisition or adoption of a new weapon, means or method of warfare' its use would 'in some or all circumstances be prohibited by international law'. This mechanism is colloquially referred to as a 'weapon review', 'legal review' or 'Article 36 review'. The importance of conducting Article 36 reviews is widely recognized and is increasingly stressed in the light of ongoing developments in civilian and military technology. The conduct of Article 36 reviews is essential to determining whether the adoption of new technologies might cause any significant concern from a humanitarian perspective and ensuring that states' armed forces are capable of conducting hostilities in accordance with their international obligations.

This SIPRI report presents the authors' key takeaways from a conference convened by SIPRI to discuss the importance of, and challenges associated with, reviewing the legality of weapons, means and methods of warfare that are based on three emerging fields of technology: cyberwarfare technologies, artificial intelligence and robotics, and human enhancement. It establishes that although these three technology areas are at various levels of maturity (from mature to still emerging and experimental), it is beyond dispute that they will have a dramatic impact on the future of warfare as they all have the potential to fundamentally change the way force is applied, and critical decisions are made, on the battlefield. The report also finds that despite their technical and operational differences, the military applications derived from these technology areas raise similar challenges as far as the conduct of Article 36 reviews is concerned.

## I. Cross-cutting challenges

The first challenge is that such technologies require a re-examination of old legal concepts. This is particularly obvious in the case of cyberwarfare technologies, which invite a rethink of what have been unequivocal concepts in international law, such as 'weapon', 'attack' or 'armed conflict'.

Second, the legal experts conducting the Article 36 review must now have acquired expertise on a much broader range of technologies. They must have a good grasp of computer science, robotics, biotechnology and neuroscience. They may not need to be experts but in order to do their job properly, they do need sufficient understanding of the underlying technologies in the weapons, means and methods of warfare they are to review. This includes being able to help technical personnel translate legal requirements into engineering decisions, and understanding the results of tests and evaluations.

Third, new methodologies are required for the assessment of performance and the risks posed. Testing of cyberweapons, autonomous weapons or methods of human enhancement faces new and in some cases very challenging requirements. In the case of autonomous weapon systems, existing methods of testing and evaluation do not yet allow an assessment of the performance and reliability of complex learning systems.

## II. Recommendations

The report outlines some important mechanisms for identifying concrete and viable solutions to the challenges to the review process posed by emerging technologies.

### Build on existing elements of best practice for the conduct of Article 36 reviews

Reviewing authorities should look to build on existing elements of best practice already identified by the International Committee of the Red Cross. This could be done in the following ways.

1. Start the review process as early as possible and incorporate it into the procurement process at key decision points.
2. Provide military lawyers involved in the review process with additional technical training. Engineers and systems developers should also be informed about the requirements of international law so that they can factor these into the design of the weapons and means of warfare.
3. Involve all relevant stakeholders (e.g. lawyers, systems developers, testers and end users) in the test and evaluation phases.
4. Rely on computer simulations when possible to reduce the cost of the procedure.

### Strengthen transparency and cooperation in the area of Article 36 reviews

Increased transparency on Article 36 review procedures could become a virtuous circle in at least the following three ways.

1. It would allow states that conduct reviews to publicly demonstrate their commitment to legal compliance.
2. It would be of assistance to states that are seeking to set up and improve their weapon review mechanisms and thereby create the conditions for more widespread and robust compliance.
3. It could facilitate the identification of elements of best practice and interpretative points of guidance for the implementation of legal reviews, which would strengthen international confidence in such mechanisms.

Cooperation is also an effective way to address some the outstanding conceptual and technical issues raised by emerging technologies. Dialogues, expert meetings and conferences can allow generic issues to be debated and addressed in a manner that does not threaten the national security of any state.

### Support targeted research

Finally, the above-mentioned challenges cannot be overcome without relevant and rigorous research. The following general outstanding issues deserve the particular attention of scholars and practitioners.

1. *Cyberwarfare technologies*. How to apply the concepts of distinction, proportionality and precaution in an environment that is primarily characterized by dual-use technology.

2. *Artificial intelligence and robotics*. How to verify the predictability of autonomous weapon systems' compliance with international law.

3. *Human enhancement*. At what point does an enhanced soldier cease to be a human being and become a mere weapon for the purpose of an Article 36 review?

# Abbreviations

| | |
|---|---|
| DDOS | Distributed denial of service |
| DOS attack | Denial-of-service attack |
| ICRC | International Committee of the Red Cross |
| IHL | International humanitarian law |
| IHRL | International human rights law |
| MHE | Military human enhancement |
| R&D | Research and development |
| RAT | Remote access tool |

# 1. Introduction

The right of states to choose the means and methods of warfare is not unlimited. International law includes both general rules and treaty law that prohibit or restrict specific effects, types of weapons or means and methods of warfare in armed conflicts. As a general rule, international humanitarian law (IHL) prohibits the use of weapons, means and methods of warfare that cause superfluous injury or unnecessary suffering, or damage military objectives and civilians or civilian objects without distinction. There are also a number of rules under treaty and customary law that ban specific types of weapons (e.g. biological and chemical weapons or blinding laser weapons) or restrict the way in which they are used, such as the 1907 Convention Relative to the Laying of Automatic Submarine Contact Mines.[1] These restrictions and prohibitions are intended to set minimum standards of humanity during armed conflicts.

As a complement to, and reinforcement of, these limitations, international law—specifically Article 36 of the 1977 Additional Protocol to the 1949 Geneva Conventions (Additional Protocol I)—imposes a practical obligation on states to determine whether 'in the study, development, acquisition or adoption of a new weapon, means or method of warfare' its use would 'in some or all circumstances be prohibited by ... international law'.[2] This mechanism is colloquially referred to as a 'weapon review', 'legal review' or 'Article 36 review'. The importance of conducting Article 36 reviews is widely recognized and is increasingly stressed in the light of ongoing developments in civilian and military technology. The conduct of Article 36 reviews is essential to determining whether the adoption of new technologies might cause any significant concern from a humanitarian perspective and ensuring that states' armed forces are capable of conducting hostilities in accordance with their international obligations.

The ability of Article 36 reviews to control technological developments can, however, be undermined by a number of issues. The first, and perhaps most fundamental, is lack of compliance. Only a small number of states are known to have a formal Article 36 review mechanism in place. Moreover, Article 36 does not provide any concrete guidance about how states should formalize such a process. The legal review mechanism may therefore differ from country to country in terms of format, method of working, mandate and level of authority. States may also be unequally equipped to comply with the obligation set out in Article 36. Some might not have the practical experience, financial resources or expertise to conduct Article 36 reviews, particularly as new technologies make the review process increasingly complex and expensive.

The lack of widespread compliance and the difference in how states conduct reviews, combined with the challenges that new technologies pose for the review process, are sources of concern for many of the civil society organizations that work on arms control and humanitarian issues. Specifically, they do not trust Article 36 to be sufficient to prevent the development or use of weapons that violate international law, and suggest that a new regulation or prohibition might be needed to control certain types of emerging military applications, such as lethal autonomous weapon systems.

SIPRI recognizes that technological change presents major challenges for the conduct of Article 36 reviews and also that this mechanism is currently the only one the international community has in place to prevent the development and use of new types of weapons, means and methods of warfare that might violate current or future international law. SIPRI is therefore committed to supporting more widespread and robust

[1] Convention Relative to the Laying of Automatic Submarine Contact Mines, opened for signature 18 Oct. 1907, entered into force 26 Jan. 1910.

[2] 1977 Protocol I Additional to the Geneva Conventions, and Relating to the Protection of Victims of International Armed Conflicts, opened for signature 12 Dec. 1977, entered into force 7 Dec. 1978.

compliance with the requirements of Article 36. SIPRI's most recent contribution in this regard was the organization of an international conference in Stockholm on 20–21 September 2017, where 75 experts from government and civil society had an opportunity to discuss the range of challenges associated with the review of weapons, means and methods of warfare based on three key emerging technologies: cyber-warfare technologies, artificial intelligence and robotics, and human enhancement.

This SIPRI report presents the authors' key takeaways from the conference. It aims to provide practitioners and experts from government and civil society with an opportunity to learn about the state of these technologies and their potential impact on the conduct of warfare, as well as the legal, operational and technical issues that these technologies raise for the Article 36 review process. It also seeks to suggest concrete solutions that might facilitate more widespread, robust and sustainable compliance with the requirements of Article 36 of Additional Protocol I.

Chapter 2 provides a brief introduction to Article 36 reviews. Chapters 3, 4 and 5 discuss the three key emerging technologies, and chapter 6 presents the conclusions.

# 2. Article 36 reviews: what they are and why they matter

## I. Article 36 and its requirements

Article 36 of Additional Protocol I states that

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

### Terms of reference

Article 36 specifically refers to 'a new weapon, means or method of warfare'. These terms are not defined in Additional Protocol I, however, and are therefore subject to different interpretations. McClelland notes that the meaning of the term 'weapon' is fairly straightforward, as it 'connotes an offensive capability that can be applied to a military object or enemy combatant'.[1] There are, however, different understandings of what types of weapon are covered. For the International Committee of the Red Cross (ICRC) and most of the states that conduct weapon reviews, the term refers to 'weapons of all types—be they anti-personnel or anti-materiel, "lethal" or "non-lethal"—weapons systems'.[2] When ratifying Additional Protocol I, however, some states added reservations to exempt certain types of weapon from the scope of application of Article 36. Germany, for example, applies the rules introduced by Additional Protocol I exclusively to conventional weapons.[3] Equipment of a dual-use nature is not subject to review, unless it can be determined that it directly contributes to the conduct of warfare.

What constitutes a 'means of warfare' is more difficult to determine. According to McClelland, the concept refers to military equipment that is not a weapon per se but 'nonetheless has an impact on the offensive capability of the force to which it belongs'.[4] To assess whether a piece of military equipment counts as a means of warfare, it is therefore necessary to understand how it works and how it may be used on the battlefield. For instance, a surveillance system will be subject to a review if it can be established that it collects and processes information used in the targeting process.

The terms 'means' and 'method of warfare' must therefore be read together. The ICRC guide to weapon reviews explains that it is necessary to examine not only the design and the purpose of the equipment, but also the way in which it is expected to be used on the battlefield—the method of warfare.[5] A weapon or means of warfare may be lawful or illegal depending on the manner and circumstances in which it is used. That is why Article 36 spells out the need to determine whether the employment of weapons, means and methods of warfare would 'in some or all circumstances' be prohibited by international law. It is, however, generally accepted that the examination should focus on the 'normal or expected use' of a weapon, means or method of warfare. The ICRC's commentary acknowledges that 'a state is not required to foresee

[1] McClelland, J., 'The review of weapons in accordance with Article 36 of Additional Protocol I', *International Review of the Red Cross*, vol. 85, no. 850 (June 2003), p. 404.

[2] International Committee of the Red Cross (ICRC), *A Guide to the Legal Review of Weapons, Means and Methods of Warfare* (ICRC: Geneva, 2006).

[3] According to an interpretative declaration dated 14 Feb. 1991 made by Germany on deposit of its instrument of ratification of Additional Protocol I.

[4] McClelland (note 1), p. 405.

[5] ICRC (note 2), p. 10.

or analyse all possible misuses of weapons, for almost any weapon can be misused in ways that would be prohibited'.[6]

Article 36 spells out that the requirement to review applies throughout the different phases of the procurement process: 'in the study, development and adoption of a new weapon, means or method of warfare'. Weapons acquired for the first time from another state should be subject to a review. Weapons acquired by a country before its ratification of Additional Protocol I are, in theory, excluded from the scope of application of Article 36. However, it is generally acknowledged that all modifications to the design or use of a weapon or means of warfare that might affect that weapon's capability and effect should trigger a review process.

## Review criteria

Article 36 requires states to consider the general provisions of IHL and any other international law applicable to that state, including in particular rules prohibiting specific weapons and means of warfare or restricting the method by which they can be used. Typically, the legal assessment can be broken down into three steps.[7]

Step 1 is the initial determination that a state has to make about whether the use of the weapon or means of warfare under review and the method by which it is to be used are already prohibited or restricted by a treaty to which it is a party or by customary international law.[8]

In step 2, if the weapon or means of warfare under review or the method by which it is to be used is not subject to any specific prohibition or restriction, the state must examine it in the light of the general rules found in Additional Protocol I and other treaties that bind the state, or in customary international law. These include (*a*) the prohibition on using 'projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering';[9] (*b*) the prohibition on employing weapons indiscriminately or indiscriminate means and methods of warfare, that is, using weapons, means and methods of warfare to strike military objectives and civilians or civilian objects indiscriminately;[10] and (*c*) the prohibition on using 'methods or means of warfare which are intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment'.[11]

As these prohibitions are largely context-dependent, the state conducting the review must take account of the environment in which the weapon is intended to be used. The use of a weapon may be lawful in one context but unlawful in another. Such an assessment might lead to the definition of conditions that can be integrated into the rules of engagement or operating procedures associated with the weapon.[12]

In step 3, should there be no relevant treaty or customary law, the state must consider the weapon in the light of the 'Martens Clause' and examine whether the weapon, means or method of warfare is of a nature that contravenes 'the principles of humanity' or 'the dictates of public conscience'.[13]

[6] ICRC, 'Commentary on the Additional Protocol, paragraph 1469', 1987.
[7] ICRC (note 2), p. 11.
[8] See the list by the ICRC (note 2).
[9] Article 35 of Additional Protocol I.
[10] Articles 48, 51 and 55 of Additional Protocol I.
[11] Articles 35 and 55 of Additional Protocol I.
[12] ICRC (note 2), p. 15.
[13] This clause is found in IHL treaties dating back to 1899; its primary modern incarnation is in Article 1(2) of Additional Protocol I, which states as follows: 'In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from dictates of public conscience.' The interpretation and application of the Martens Clause is a matter of debate: some consider that it imposes a test for the lawfulness of new weapons, while others believe it provides guidelines for the evolution of customary or treaty law.

It remains a matter of debate whether a state should give consideration to international human rights law (IHRL) in the review process, as there are different views on whether IHL displaces IHRL entirely in the area of armed conflict, or IHL and IHRL are complementary and both apply during armed conflict.[14] Some states, such as Sweden, Switzerland and the United Kingdom, see some value in considering IHRL in any weapon review because military personnel could in some situations, such as peacekeeping, use the weapon to conduct law enforcement missions.[15] The fundamental rights most relevant to an Article 36 review are (*a*) the right to life, which prohibits a state from arbitrarily depriving a person of his or her life; (*b*) the right to freedom from torture and other forms of cruel, inhuman or degrading treatment; and (*c*) the right to health, understood as the right to the enjoyment of the highest attainable standards of physical and mental health.[16]

Although it is not required by Article 36, some states, such as Sweden or the UK, consider it useful to give consideration to future development of the law, as this would avoid the consequences of approving and procuring a weapon, means or method of warfare that is likely to be restricted or prohibited in the future.[17]

## Conducting legal reviews

The ICRC's guide to legal reviews of weapons, means and methods of warfare notes that 'Article 36 of Additional Protocol I does not specify how a determination of the legality of weapons, means and methods of warfare is to be carried out'. It implies an obligation on states to establish internal procedures but does not provide any details on how these should be arranged. Consequently, the legal review mechanism often differs from country to country in terms of format, method of working, mandate and level of authority.

For the ICRC and a number of scholars who have worked on the topic, this is not necessarily problematic—for two reasons. First, from a practical standpoint, there cannot be a single model of compliance.[18] States have different needs as well as access to different human and financial resources for conducting Article 36 reviews. According to McClelland, imposing a uniform system would undermine the ability of a state to integrate the legal review process into its own weapon acquisition process.[19] Each state should, according to this argument, be able to determine what type of review mechanism is best suited to its needs. In the words of Parks, 'establishing and maintaining a weapon review programme is more important than the form it takes'.[20] Second, the fact that Article 36 does not prescribe exactly how the review process should be conducted allows states to adapt to change, notably technological change. This is discussed further below.

Ticehurts, R., 'The Martens Clause and the laws of armed conflict', *International Review of the Red Cross*, no. 317 (Apr. 1997).

[14] Hathaway, O. et al., 'Which law governs during armed conflict? The relationship between international humanitarian law and human rights law in armed conflicts', *Minnesota Law Review*, vol. 96 (2012), pp. 1883–1944.

[15] The 1990 UN Basic Principles encourage states to review less lethal weapons used for law enforcement purposes: 'Government and law enforcement agencies should develop a range of means as broad as possible and equip law enforcement officials with various types of weapons and ammunition that would allow for a differentiated use of force and firearms.' For more commentary see Casey-Maslen, S., Corney, N. and Dymond-Bass, A., 'The review of weapons under international humanitarian law and human rights law', ed. S. Casey-Maslen, *Weapons Under International Human Rights Law* (Cambridge University Press: Cambridge, 2014).

[16] Casey-Maslen, Corney and Dymond-Bass (note 15).

[17] ICRC (note 2), p. 11.

[18] Hays Parks, W., 'Conventional weapons and weapons reviews', *Yearbook of International Humanitarian Law*, vol. 8 (2005), p. 107; and McClelland (note 1), p. 414.

[19] McClelland (note 1), p. 414.

[20] Hays Parks (note 18), p. 107.

While one size cannot fit all in the area of Article 36 reviews, it can nonetheless be useful to identify elements of best practice that might help states set up or reform their own weapon review mechanisms. For this reason, the ICRC produced its own guide in 2006.[21] Drawing on existing practice, it suggests the types of weapons that should be subject to legal review, a legal framework for the review of new weapons, means and methods of warfare, and the types of empirical data to be considered by the review. It makes suggestions about how the review mechanism should be established, structured and composed, and under whose authority it might be placed. It also describes how a review body might operate and take decisions.

Practitioners from various countries have also made recommendations on how to conduct or improve weapon review procedures.[22] Their key recommendations are as follows: (*a*) start the review process as early as possible in the procurement process, and if possible incorporate legal reviews into the acquisition process at key decision points; (*b*) take a multidisciplinary approach, seeking input from various fields of expertise (legal, technical, operational and medical); and (*c*) examine the empirical evidence provided by the manufacturer and intended end user, and if necessary conduct tests and evaluations to assess the weapon's performance and the possible risks associated with its use.

## II. Why Article 36 matters: technological changes and the conduct of warfare

The importance of conducting weapon reviews is widely recognized and being increasingly stressed in the light of the fast pace of innovation in the field of information technology, in areas such as advanced computing and communications, nanotechnology and synthetic biotechnology. Advances in science and technology could result in weapon and equipment developments that transform the conduct of modern warfare. Article 36 reviews are currently the only binding mechanism the international community has to force states to assess whether the employment of these weapons, means and methods, which build on fundamentally new technologies, raises any significant concerns from a humanitarian perspective. Most notably, the question is whether they cause unnecessary suffering or superfluous injury, or if they target military objectives, civilians and civilian objects indiscriminately.

However, the novelty of the technology can in some cases make the process of conducting an Article 36 review very difficult. It might necessitate revisiting old legal concepts anew or pose new risks that may themselves require new methods of risk assessment.

Chapters 3 to 5 explore this conundrum by studying three emerging areas of military technology that are expected to fundamentally change the way in which the military uses force and makes decisions on the battlefield: cyberwarfare, autonomous systems warfare and military human enhancement technologies. Each chapter provides a brief overview of the technology and its application, and examines the legal considerations that would have to be taken into account in an Article 36 review of a weapon, means or method of warfare that builds on that technology. The concluding chapter (chapter 6) discusses the extent to which the technology poses new and potentially difficult challenges for the conduct of legal reviews.

[21] ICRC (note 2), p. 15.

[22] Hays Parks (note 18), pp. 55–142; McClelland (note 1); and Boothby, W., *Weapons and the Law of Armed Conflict* (Oxford University Press: Oxford, 2009).

# 3. Reviewing the legality of cyber weapons, means and methods of warfare

One of the most remarkable phenomena in the realm of modern warfare in recent decades has been the emergence of cyberspace as a new warfighting domain. Cyberspace is now commonly depicted as the fifth warfighting domain, along with land, sea, air and space. The conduct of military operations in this domain, however, has few similarities with the other four. The question of whether existing rules of IHL are adequate for regulating the conduct of cyber-operations has become a matter of contention among the community of international law scholars. This chapter explores the implications of this debate for the conduct of Article 36 reviews. It translates the product of academic discourse into concrete legal advice for Article 36 review practitioners and military lawyers advising commanders on the impact of international law on cyber-operations.

## I. Cyberwarfare toolbox

Cyberwarfare is an umbrella term for a spectrum of activities and there are fundamental disagreements within the community of experts about the terminology used to label and define these activities. This contention is caused partly by the technical specificities of cyberspace and partly by the fact that the categorization of military operations is fraught with major policy and legal implications. These affect, for instance, the determination of which tools and capabilities should be subject to an Article 36 review. To provide some context for the legal analysis, this section discusses the spectrum of military activities that may be conducted in cyberspace.

### Spectrum of operations

A simple and uncontroversial way of mapping the range of operations that the military can conduct in cyberspace is to classify them according to their intended effects. In this way, cyberwarfare operations can be sorted into three generic categories depending on whether they affect the availability, confidentiality or integrity of information systems, including computer devices and computer networks, and the data they hold.

*Operations targeting the availability of computer devices, networks or data*

Operations targeting the availability of information systems or their data are operations that seek to undermine access to, or the use of, these systems or the information they contain without causing physical damage or injury.[1] The commonly used method of attack is known as a denial-of-service attack (DOS attack). DOS attacks can disrupt information systems in two ways: by flooding the targeted system with a great number of simultaneous requests or by triggering errors in the systems. The former technique works primarily with systems that are connected to the Internet or a network. For the latter, the attacker would need to first gain access to the systems. (The methods for achieving this are discussed further below.) DOS attacks can be conducted from a single computer but usually rely on a large network of 'botnets', that is, computers that the attacker has secretly compromised and taken control of using malicious software, or malware. In such cases the attack is referred to as a distributed denial of service (DDOS).

[1] Brown, G. and Tullos, O., 'On the spectrum of cyberspace operations', *Small Wars Journal*, 11 Dec. 2012.

*Operations against the confidentiality of computer devices, networks or data*

Operations against the confidentiality of information systems and their data, known as access operations, seek to breach an information system either to collect information or to enable operations that seek to undermine the availability or integrity of the targeted system or a system connected to it. Depending on whether the purpose of the operation is the former or the latter, the operation may also be described as cyber-espionage or an 'enabling operation'. There are multiple methods for gaining access to systems. A common method is to trick the user of the targeted system into installing a 'Trojan horse', or malware usually disguised as legitimate software from a legitimate source, that provides the attacker with a remote access tool (RAT), or backdoor access to the targeted system. The attacker can then use the RAT to steal, or secretly manipulate, data held on the affected system. The Trojan horse can be given various malware payloads that have different properties. A key logger, for instance, is malware that records key logs for the purpose of accessing confidential data such as passwords. Payloads that enable access, and copy and export content are generally referred to as spyware. To be successful, access operations need to be stealthy: the victim must remain unaware of the security breach. The RATs used for this type of operation are usually mounted with 'rootkits', another type of program that enables the malware to disguise its activity.[2]

*Operations against the integrity of computer devices, networks or data*

Operations against the integrity of information systems or the data they hold are operations that involve manipulating or altering the data held on the targeted systems, the foreseeable results of which could include damage to, or the destruction of, property (e.g. data or physical infrastructure), as well as death or injury to persons.[3] This kind of malevolent operation is more sophisticated as it requires greater amounts of manpower and technical expertise, as well as more information on the target. A famous example of malware used in such an operation is *Stuxnet*, a sophisticated computer virus specifically designed to damage plutonium processing centrifuges in Iran.[4]

## Some technical considerations

The cyberwarfare toolbox has a number of technical specificities that are worthy of note for the purposes of an Article 36 review.

*Cyber-tools need to be tailor-made*

The methods and tools used to affect the availability, confidentiality or integrity of information systems and the information they hold typically need to be tailored to their targets, as they generally involve exploiting vulnerabilities that are specific to the target and the operational context. The code used to conduct the attack against a system may also have to be changed on the fly during an operation, for instance as adversaries update their defences, possibly with antivirus software.

*The scope and sophistication of operations can vary greatly*

Methods and tools can vary a great deal in terms of both sophistication and scope. At the higher end, there are 'advanced persistent threat attacks', cyberattacks that are sophisticated and extremely discriminating in nature. The attackers know exactly

[2] A Trojan horse is neither a 'computer virus' nor a 'computer worm'; it has neither the capacity to inject itself into other files (the property of a virus) nor the capacity to propagate itself (the property of a worm).

[3] Brown and Tullos (note 1).

[4] Singer, P. W. and Friedman, A., *Cybersecurity and Cyberwar: What Everyone Needs to Know* (Oxford University Press: Oxford, 2014), pp. 114–18.

what systems they are targeting and what they are trying to achieve. These attacks are also highly covert. Attackers take their time in order to go unnoticed. They can spend months or even years on reconnaissance and intelligence-gathering activities to (*a*) identify vulnerabilities; (*b*) assess the behaviour of the target; and (*c*) determine how to extract information or corrupt the target secretly.[5] At the lower end, there is off-the-shelf malware, or malware that is ready for use by virtually any type of actor (who does not require any programming skills) and that typically exploits well-known vulnerabilities in standard information systems.

*Some malware can operate autonomously*

Some malware, known as a computer worm, has the ability to replicate and spread itself without human intervention. This means that it can operate autonomously.

## II. Determining the lawfulness of cyber weapons, means and methods of warfare

### What is to be reviewed? How to apply the concept of weapons, means and methods of warfare in cyberspace

It has been officially agreed that IHL applies in cyberspace.[6] The question of how its key concepts and rules can be concretely applied, however, is hotly debated in both academic circles and the policy community. With regard to the application of Article 36, the central dilemma is how to determine what type of cyber-capability would be classified as a weapon, means or method of warfare, and therefore be subject to a review process.

*Cyberweapons*

It is generally agreed that the concept of a cyberweapon should be reserved for cyber-means of warfare that are intended to cause harm to people or to damage objects (see box 3.1) and that it therefore excludes tools that are solely designed to access information.[7]

'Object' and 'damage', however, are open to different interpretations. Are intangibles such as software systems and computer data objects? Does the notion of damage refer only to physical damage or does it also include effects such as loss of functionality, which can render objects dysfunctional without physically damaging them? Legal experts respond differently depending on how they define the concept of cyberattack (see box 3.2).

Those who define cyberattacks in the narrow sense tend to reserve the concept of cyberweapons to cyber-tools that are designed to result in tangible physical damage, or damage that would cause a loss of usability and might result in the need to

[5] Singer and Friedman (note 4), pp. 55–60.

[6] ICRC, 'International humanitarian law and the challenges of contemporary armed conflicts' (Oct. 2015), Report 32IC/15/11 for the 32nd International Conference of the Red Cross and the Red Crescent, Geneva, 8–10 Dec. 2015; and United Nations Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunication in the Context of International Security, A/68/98, 24 June 2013. See also Koh, H., 'International law in cyberspace', Speech at the USCYBERCOM Interagency Legal Conference, Fort Meade, MD, 18 Sep. 2012.

[7] Rid, T. and McBurney, P., 'Cyber-weapons', *RUSI Journal*, no. 157, p. 7. Some legal experts refute this analysis. Arimatsu argues that such a narrow definition would fail to capture the essence of cyberweapons, as most are designed to have an indirect outcome. Blake and Imburgia point out that cyber-operations that have the capacity to seriously disrupt a country's critical infrastructure (financial sector, medical services etc.) could be defined as weapons since they would arguably breach the international peace and security the United Nations Charter was designed to maintain. This approach is criticized, however, as being overly broad and unfeasible. See Arimatsu, L., 'A treaty for governing cyberweapons: potential benefits and practical limitations', eds C. Czossesk, R. Ottis and K. Ziolkowski, *2012 4th International Conference on Cyber Conflict*, Proceedings (NATO Cooperative Cyber Defence Centre of Excellence Publications: Tallinn, 2012), p. 97; and Blake, D. and Imburgia, J. S., '"Bloodless weapons"? The need to conduct legal reviews of certain capabilities and the implications of defining them as "weapons"', *Air Force Law Review* (2011), p. 161.

**Box. 3.1.** Defining 'cyberweapons'

The following definitions have been used for 'cyberweapons'.

1. 'An object designed for, and developed or obtained for, the primary purpose of killing, maiming, injuring, damaging, or destroying.'[a]
2. 'Cyber means of warfare that are used, designed, or intended to be used to cause injury to or death of persons or damage to, or destruction of, objects.'[b]
3. 'Any device or software payload intended to disrupt, deny, degrade, negate, impair or destroy adversarial computer systems, data, activities or capabilities.'[c]

[a] Brown, G. D. and Metcalf, A. O., 'Easier said than done: legal reviews of cyber weapons', *Journal of National Security Law and Policy*, vol. 7, no.1 (2014).
[b] Schmitt, M. N. and NATO Cooperative Cyber Defence Centre of Excellence (eds), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge University Press: Cambridge, 2017).
[c] US Department of the Air Force, Secretary of the Air Force, Air Force Instruction 51-402, Legal Reviews of Weapons and Cyber Capabilities, 27 July 2011.

replace physical components of the computer system.[8] According to this view, cyber-capabilities that only damage data do not qualify as cyberweapons.[9] Experts who use a wider interpretation of cyberattack, by contrast, believe that cyber-tools that can erase or disrupt data qualify as cyberweapons if the targeted computer would require 'data restoration' in order to become functional again.[10]

### *Cyber-means of warfare*

Although there are slight differences, existing definitions of cyberweapons are sufficiently straightforward to allow a reviewer to determine whether a given cyber-tool would have to be classified as a weapon and therefore be subject to an Article 36 review. The task becomes more intricate in the case of cyber-means of warfare. As explained above, means of warfare can be understood as any type of military equipment that has an impact on the offensive capability of the force to which it belongs. In cyberspace, this definition creates a requirement to review any cyber-related device, materiel, instrument, mechanism, equipment or software designed for, or used (or intended for use) in, an attack.[11] This would include any type of botnet, malware or spyware used to gain access to a computer to prepare or support an attack.

### *Cyber-methods of warfare*

Cyber-methods are generally understood as the cyber-tactics, techniques and procedures by which hostilities are conducted. It is worth noting that the notion of hostilities covers more types of operations than just those that constitute a cyberattack. According to this definition, a botnet is the means of warfare and the denial of service the method.[12]

[8] This is the view of Schmitt and the majority of experts who contributed to the *Tallinn Manual on the International Law Applicable to Cyber Warfare*. See Schmitt, M. N. (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge University Press: Cambridge, 2013), Rule 30(11). It is worth mentioning that the notion of damage applies not only to the targeted computer (first order of effect), but also to the impact on any potential installation it might service or control (second order of effect) as well as the impact on the people affected by the shutdown of the services of that installation (third order of effect). Boothby, W. H., 'Methods and means of cyber warfare', *International Law Studies*, vol. 89 (2013), p. 389.

[9] Schmitt, M., '"Attack" as a term of art in international law: the cyber operation context', eds C. Czossesk, R. Ottis and K. Ziolkowski, *2012 4th International Conference on Cyber Conflict*, Proceedings (NATO Cooperative Cyber Defence Centre of Excellence Publications: Tallinn, 2012), p. 291.

[10] This position was envisaged by a minority of the *Tallinn Manual on the International Law Applicable to Cyber Warfare*'s experts and constitutes the official position of the ICRC. It derives from the idea that limiting the interpretation of damage to physical effects would rule out an excessive amount of malevolent operations that could gravely affect civilian computer networks. ICRC (note 6); and Droege, C., 'Get off my cloud: cyber warfare, international humanitarian law, and the protection of civilians', *International Review of the Red Cross*, vol. 94, no. 886 (June 2012).

[11] Schmitt (note 8).

[12] Schmitt, M. N. and NATO Cooperative Cyber Defence Centre of Excellence (eds), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge University Press: Cambridge, 2017) pp. 452–53.

**Box 3.2.** Legal definition of 'cyberattack'

There is an ongoing debate among legal experts over the definition of the term 'cyberattack'. The way the term is legally defined greatly influences the protection that international humanitarian law affords civilian infrastructure. It also affects the scope of the Article 36 review process, as it determines the type of cyber-capabilities that are deemed relevant to the conduct of an attack. There are three general interpretations:

1. The narrow interpretation is that the term 'cyberattack' applies only to an operation that causes violence to persons or physical damage to objects.[a]
2. A second more extensive view makes the analysis dependent on the action required to restore the functionality of the object, network or system.[b]
3. A third approach focuses on the effects that the operation has on the functionality of the system. According to this view, an operation aimed at impairing the functionality of a system to 'neutralize' it, without necessarily causing physical damage, still amounts to an attack.[c]

Any operation expected to cause death, injury or damage constitutes an attack, including where such harm is due to the foreseeable indirect or 'reverberating effects' of an attack such as the death of a patient in intensive care caused by a cyberattack against the electricity supply to a hospital.

[a] Gill, T. D., Heinsch, R. and Geiss, R., 'The conduct of hostilities and international humanitarian law: challenges of 21st century warfare', International Law Association Study Group Interim Report (2014); and Schmitt, M. N. (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge University Press: Cambridge, 2013).

[b] Schmitt, M., '"Attack" as a term of art in international law: the cyber operation context', eds C. Czossesk, R. Ottis and K. Ziolkowski, *2012 4th International Conference on Cyber Conflict*, Proceedings (NATO Cooperative Cyber Defence Centre of Excellence Publications: Tallinn, 2012), p. 291; and Boothby, B., 'How will weapons reviews address the challenges posed by new technologies?', *Military Law and the Law of War Review*, vol. 52, no. 1 (2013).

[c] ICRC, 'International humanitarian law and the challenges of contemporary armed conflicts', Oct. 2015, Report 32IC/15/11 for the 32nd International Conference of the Red Cross and the Red Crescent, Geneva, 8–10 Dec. 2015, p. 39.

## Rules and principles of international law applicable to the review of cyber weapons, means and methods of warfare

There are no treaties that specifically restrict or ban the acquisition, possession or use of cyber weapons, means or methods of warfare.[13] Cyber-operations are therefore regulated by the general rules and principles of international law. The key rules and principles to consider are the prohibition on indiscriminate weapons, means and methods of warfare, the prohibition on perfidious acts and the relevant aspects of the law of neutrality.

### *The prohibition on the employment of indiscriminate weapons, means and methods of warfare*

Given the inherent interconnectivity and the dual-use nature of cyberspace, the prohibition on indiscriminate weapons, means and methods of warfare is perhaps the most crucial of the general rules of international law to consider in the conduct of an Article 36 review.[14] It is generally understood that a cyberweapon might easily spread out of control to civilian networks and should therefore be deemed an inherently indiscriminate weapon.[15] According to Boothby, 'whether the cyberweapon is indiscriminate by nature will depend on whether its effects can be limited to a computer node, network or system that is the military objective the attack is intended to engage on and whether it will only attack that military objective'.[16] There are therefore two

[13] Lin, H., 'Governance of information technology and cyber weapons', ed. E. D. Harris, *Governance of Dual-use Technologies: Theory and Practice* (American Academy of Arts and Sciences: Cambridge, MA, 2016), p. 124.

[14] Military communications almost always transit through civilian-owned and -operated networks. Talbot Jensen, E., 'Cyber warfare and precautions against the effects of attacks', *Texas Law Review*, vol. 88 (June 2010), p. 1542.

[15] ICRC (note 6).

[16] Boothby, W., 'How will weapons reviews address the challenges posed by new technologies?', *Military Law and the Law of War Review*, vol. 52, no. 1 (2013).

risk scenarios that the reviewing authority should consider in particular: (*a*) the risk that the cyberweapon under review might affect or damage civilian systems and military networks due to the entanglement of civilian networks in military networks;[17] and (*b*) the risk that a cyberweapon might have the ability to replicate itself (a property of computer worms) and spread itself indiscriminately.[18]

It should be noted that the way in which the reviewing authority interprets the notion of cyberattack will determine whether civilian data, in particular data belonging to certain categories of object, would be protected by this general rule and should therefore be included in the potential effects on civilian data as part of the review process. The ICRC considers that the obligation to respect and protect civilian facilities also extends to data belonging to those facilities and therefore that a cyber-capability that could damage or impair data belonging to a civilian network's infrastructure should be deemed unlawful.[19]

A reviewing authority must also take into account in its assessment the possible reverberating or indirect effects of the attack.[20] This means that it should consider the direct effects on the targeted systems (the first order of effect) and the impact on any potential installation it might service or control (the second order of effect), as well as the impact on the people directly affected by the loss of usability and function of the systems (the third order of effect).[21] This assessment of reverberating effects is essential to determining whether the use of the new weapon, means or method of warfare might be expected to cause superfluous injury or unnecessary suffering, or widespread and long-term damage to the natural environment.[22]

*Prohibition on perfidy*

Cyberwarfare might also require consideration of other rules of international law that are less common in a weapon review, such as the prohibition on killing or injuring an adversary through resort to perfidy.[23]

'Perfidy' is defined as 'acts that invite the confidence of the adversary to believe that he or she is entitled to, or obliged to accord, protection under IHL, with intent to betray that confidence'.[24] Perfidy is forbidden if it leads to—or, according to some experts, if it is intended to lead to—the injury, death or capture of an adversary.[25]

Some legal experts argue that the notion of 'confidence' could be applied to a cyber-system, meaning that a cyber-capability designed to invite the confidence of an adversary's computer, such as a Trojan horse, might be judged unlawful if the end use

[17] Geiss, R. and Lahmann, H., 'Cyber warfare: applying the principle of distinction in an interconnected space', *Israel Law Review*, vol. 45, no. 3 (Nov. 2012), p. 383.

[18] Droege (note 10), pp. 553–56; Schmitt, M., 'Wired warfare: computer network attack and *jus in bello*', *International Review of the Red Cross*, vol. 84, no. 846 (June 2002), p. 389; and Dinstein, Y., 'The principle of distinction and cyber war in international armed conflicts', *Journal of Conflict and Security Law*, vol. 17, no. 2 (July 2012), p. 264.

[19] ICRC (note 6).

[20] Boothby (note 8), p. 389.

[21] The question of whether a reviewing authority must take account of subsequent orders of effect is debatable, not least given that it might be difficult to attribute and estimate them. Boothby (note 8), p. 389.

[22] Blount, P. J., 'The preoperational legal review of cyber capabilities: ensuring the legality of cyber weapons', *Northern Kentucky Law Review*, vol. 39, no. 2 (2012), pp. 217–20; and Harrison Dinniss, H., *Cyber Warfare and the Laws of War* (Cambridge University Press: New York, 2012), pp. 221–27.

[23] Rowe, N. C., 'Perfidy in cyberwarfare', eds F. Allhoff, N. G. Evans and A. Henschke, *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-first Century* (Routledge: Abingdon, 2013), pp. 394–99; and Greer, M. J., 'Redefining perfidy', *Georgetown Journal of International Law*, vol. 47, no. 1 (2015), pp. 270–73.

[24] The *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* notes that 'in order to breach the prohibition against perfidy, the perfidious act must be the proximate cause of death or injury'. It further points out that 'proximate cause should not be confused with temporal proximity. It is possible that the perfidious act inviting the adversary's confidence will occur at a point in time that is remote from the act that causes death or injury'. Schmitt and NATO Cooperative Cyber Defence Centre of Excellence (note 12), pp. 491–92.

[25] Henderson, I., den Dulk, J. and Lewis, A., 'Emerging technology and perfidy in armed conflict', *International Law Studies*, vol. 91 (2015), pp. 483–85.

is to cause injury or death.[26] According to the *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (*Tallinn Manual*), one example would be the use of malware to hack into the pacemaker of an enemy commander to allow the attacker to remotely cause a heart attack.[27]

The perfidy rule arguably reinforces the need to interpret the obligations in Article 36 in the broadest sense and review not only the tools that qualify under the definition of cyberweapons, but also means of cyberwarfare such as access tools, as these can be of critical importance in the conduct of an attack.

*The law of neutrality*

The law of neutrality is generally irrelevant to the conduct of Article 36 reviews. However, given the global and interconnected nature of cyberspace, its rules need to be taken into consideration in an Article 36 review of a cyber weapon or means or method of warfare. Under the law of neutrality, belligerents are forbidden to move troops or convoys of either munitions or supplies across the territory of a neutral power (Article 2 of the 1907 Hague Convention).[28] Article 8 of the 1907 Hague Convention, however, states that 'a neutral Power is not called upon to forbid or restrict the use on behalf of the belligerents of telegraph or telephone cables or of wireless telegraphy apparatus belonging to it or to companies or private individuals'. The question of whether a cyberweapon may be transferred through the cyber-infrastructure of a neutral country is therefore debatable. The majority of experts who contributed to the *Tallinn Manual* were of the view that it is prohibited under Article 2. It was also generally agreed that Article 2 would prohibit the use of botnets located on neutral territory.[29]

*Taking control of an enemy weapon*

One situation that is unique to cyberwarfare would be a case in which a cyber-capability is designed to take control of an enemy's weapon system in order to use it against that enemy. In such cases, the review of that capability should take into consideration not only the means of taking control, but also the enemy's system, as the enemy system itself could be deemed unlawful (e.g. if it was a prohibited chemical weapon).

## Checklist for the review of cyber weapons, means and methods of warfare

It should be stressed that a review of a cyber-capability cannot properly take place without a preliminary clarification of the state's position on a number of contentious legal issues, starting with the definition of cyberattack, which is essential in order to delimit the material scope of the review. The reviewing authority should also clarify (*a*) whether the idea of a cyberattack includes operations intended to reduce the functionality of a system or damage data belonging to that system without causing physical harm or damage; and (*b*) how many orders of effect should be considered in any assessment of an attack.

The key legal questions that a reviewing authority needs to consider in any review of a cyber-capability can be summarized as follows.

[26] The perfidy rule does not extend to perfidious acts that result in damage to or destruction of property.

[27] Schmitt and NATO Cooperative Cyber Defence Centre of Excellence (note 12), p. 455.

[28] Convention (IV) respecting the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land (Hague Convention), adopted 18 Oct. 1907, entered into force 26 Jan. 1910.

[29] Schmitt and NATO Cooperative Cyber Defence Centre of Excellence (note 12).

1. Are the effects of the cyber weapon, means or method of warfare intended to cause physical damage or loss of functionality, or to allow access to, and control of, data?
2. Are the intended effects indiscriminate? Given the interconnectivity of information technology, is there a risk that the operation might have unintended effects on civilians and civilian objects, possibly including civilian data?
3. Are the intended effects of a nature that might directly or indirectly cause superfluous injury or unnecessary suffering, or long-term, widespread and severe damage to the natural environment?
4. If the intended effect is to gain access to an information system, will that access be used for spying or in operations that might be deemed perfidious under IHL?
5. Would the cyber-operation involve transferring data through cyber-infrastructure or exploiting computers located in a neutral territory?
6. Is the cyber-operation a means for taking control of an enemy's weapon systems? If so, the characteristics of the enemy system will have to be reviewed as well.

Depending on the response to these questions, the reviewing authority could place restrictions or make recommendations on the conduct of the cyber-capability. These could be integrated into the programming of the systems, the rules of engagement and training programmes. Such restrictions and recommendations might include the following.

1. Restrictions on the conditions of use, such as strictly defining the target and the network through which the cyberweapon can be transmitted.
2. A recommendation that any malware used as part of a cyber-capability should not be able to replicate itself, should include a kill switch and should automatically self-destruct once the objective of the operation has been achieved.
3. A recommendation to undertake a preventive network mapping before launching a cyber-operation in order to improve compliance with the law of targeting, particularly the rules on distinction, proportionality and precaution in attack.

## Outstanding issues

It should be acknowledged that addressing the above-mentioned questions will be difficult in practice. There are many concrete challenges linked to the nature of the technology, which affect the feasibility of the review process. These include determining the timing of the review, the appropriate testing and evaluation process, and the relevant review methodology.

### *Timing of the review*

The computer code behind a cyber-capability often needs to be updated, sometimes even during an operation itself. This would be the case, for instance, if an adversary were to update its defences with a new firewall or antivirus software, or patch the vulnerability that the cyber-capability is exploiting. From the perspective of an Article 36 review, this raises a number of problems about when the review should be conducted, the version of the code that should be reviewed and the types of modification that might be deemed to trigger a new review.

### *Testing and evaluation*

A further challenge is related to the obligation to generate empirical evidence of the effect of a new weapon, means or method of warfare through testing and evaluation.

How can damage be quantified or qualitatively assessed when the cyber weapon, means or method of warfare, and probably also the target, are entirely made of computer code and therefore intangible? Some countries use 'cyber-ranges' to test new cyber-capabilities. Cyber-ranges are virtual test beds where the interaction between attackers and defenders can be simulated in a controlled environment.[30] These are highly complex simulations but, as with any type of simulation, they are built on a simplified representation of the real world, and might therefore miss crucial elements that could potentially affect the legality of a weapon.[31] It is difficult to capture in a simulation the reverberating effects of a cyber-operation, beyond the second and third order of effect, particularly when the targeted systems are connected to the Internet or are controlling systems that have an effect on the physical world (i.e. the supervisory control and data acquisition systems of critical infrastructure).[32]

*Changing methodology?*

Experts are divided as to how the community of Article 36 review practitioners should deal with these practical difficulties. Some argue that these difficulties reinforce the need to implement Article 36 reviews as a process that runs throughout the lifecycle of a weapon or means of warfare. This would require not just one but multiple reviews. Some experts have suggested a more pragmatic approach, noting that it might be more practical and effective to review the operation rather than the capabilities themselves; that is, to shift the focus from the Article 36 review to the operational legal review. Brown and Metcalf argue that an operational legal review would still address international law concerns and review important aspects of the Article 36 review checklist.[33] It would also give the reviewer a better grasp of the context of the operation and therefore a clearer understanding of the capabilities and how they might be used.

[30] Brangetto, P., Çalişkan, E. and Rõigas, H., *Cyber Red Teaming: Organisational, Technical and Legal Implications in a Military Context* (NATO Cooperative Cyber Defence Centre of Excellence Publications: Tallinn, 2015); and Grant, T., 'Specifying functional requirements for simulating professional offensive cyber operations', Paper presented at the 10th International Conference on Cyber Warfare and Security (ICCWS 2015), Krueger National Park, South Africa, 24 Mar. 2015.

[31] Christensen, P., 'Test Resource Management Center and the National Cyber Range', Paper presented at the 32nd Annual National Test and Evaluation Conference, San Diego, 7 Mar. 2017.

[32] NATO Modelling and Simulation Group MSG-121, Modelling and Simulation Support for Cyber Defence 1.0, 'Technical evaluation report', STO-MP-MSG-121, 11 Apr. 2014; and Davis, J. and Magrath, S., 'A survey of cyber ranges and testbeds', DSTO-GD-0771, Oct. 2013.

[33] Brown, G. D. and Metcalf, A. O., 'Easier said than done: legal reviews of cyber weapons', *Journal of National Security Law and Policy*, vol. 7, no. 1 (Feb. 2014).

# 4. Reviewing the legality of weapons, means and methods of warfare with autonomous capabilities

Artificial intelligence and robotics have made great strides in the past three decades. One major outcome of innovation in these fields has been the remarkable progress of autonomy in weapon systems and the networks in which they are embedded. The advance of autonomy is a notable technological development in the sense that it fundamentally changes the way the military can field forces and make decisions, lethal or otherwise, on the battlefield. This chapter explores the implications of this development for the conduct of Article 36 reviews.

## I. Autonomy in weapons, means and methods of warfare

As a prelude, it might be useful to define autonomy and examine the impact that advances in autonomy in weapons, and means and methods of warfare could have on the conduct of war.

### Defining autonomy

In simple terms, 'autonomy' can be defined as the ability of a machine to execute a task, or tasks, without human input, using interactions between computer programming (perception and control algorithms) and the environment.[1] An autonomous system is, by extension, usually understood as a system—whether hardware or software—that, once activated, can perform some tasks or functions on its own. From the perspective of an Article 36 review, it is essential to note that autonomy is a relative notion, and that the level at which a system may be deemed autonomous can be analysed from three different and independent perspectives: (*a*) based on the nature of the human–machine command-and-control relationship; (*b*) based on the system's decision-making capability; and (*c*) based on the number and types of functions that are automated.[2]

*Human–machine command-and-control relationship*

When assessing a system's level of autonomy, the first dimension to consider is the extent to which humans are involved in the execution of the task carried out by the system. Using this approach, autonomous systems are often classified as (*a*) semi-autonomous (i.e. the system performs some operations autonomously, but remains under the active control of a human operator); (*b*) human-supervised autonomous (i.e. the system operates completely autonomously, but remains under the oversight of a human operator who can supervise); or (*c*) unsupervised autonomous (i.e. the system operates fully autonomously, without the direct oversight of a human operator).

*A system's decision-making capability*

A second dimension to take into consideration is the sophistication of the control algorithms, which themselves determine the extent to which the system can exercise control over its own behaviour and deal with uncertainties in its operating environment.

[1] This definition is based on one previously proposed by Williams. Williams, A., 'Defining autonomy in systems: challenges and solutions', eds A. P. Williams and P. D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (NATO Headquarters Allied Command: Norfolk, VA, 2015).

[2] Scharre, P., 'The opportunity and challenge of autonomous systems', eds Williams and Scharre (note 1).

Using this approach, autonomous systems can be sorted into the following three categories.[3]

1. *Reactive systems*. The system follows condition–action rules (also known as 'if–then rules') that explicitly prescribe how the system should react to a given sensory input. The system simply goes through a series of pre-scripted actions, so its behaviour is predictable if the rules are known.

2. *Deliberative systems*. The system uses a model of the world (information on how the world works and the reactions to the system's actions), a value function, which provides information about the desired goal, and a set of potential rules that helps it to search and plan for how to achieve the goal. To make a decision, a deliberative control system weighs the consequences of possible actions to find the most suitable actions to achieve its goal. In contrast to reactive systems, the behaviour of deliberative systems may not be entirely predictable: the overall activity of an autonomous unmanned aircraft will be predictable but individual actions may not be.

3. *Learning systems*. Learning systems can improve their performance over time through experience. They learn by abstracting statistical relationships from data. The knowledge gained serves to automatically re-parameterize and partially reprogram the system. If the system learns on the job, the behaviour of the system could become highly unpredictable unless the learning parameters (input data) are well known and understandable to an operator.

*The types of decision or function being made autonomous*

The third and perhaps most important dimension to consider is the types of decisions and functions that are being made autonomous within a system. Autonomy can serve very different capabilities in systems, including the following: (*a*) mobility (e.g motion control, navigation or take-off and landing); (*b*) targeting (e.g. target detection, classification, selection, tracking and engagement); (*c*) intelligence (e.g. detecting explosives and the location of gunfire or objects of interest); (*d*) interoperability (e.g. information sharing and collaboration with other systems); and (*e*) health management of systems (e.g. fault detection or self-refuelling).[4]

For each of these capabilities the parameters of autonomy, whether in terms of the human–machine command-and-control relationship or the sophistication of the decision algorithm, may vary greatly, even over the duration of a mission. The legal implications are also fundamentally different. The use of autonomy for targeting raises completely different issues from the use of autonomy for mobility, which is of crucial importance from the perspective of an Article 36 review.

## Autonomy and the conduct of warfare

Most of the legal issues raised by the development of autonomy in weapons, means and methods of warfare derive from the fact that autonomy has the potential to change the conduct of war in two significant ways. It allows the military to field force in new and potentially different ways and, more importantly, it affects the way people make lethal decisions on the battlefield.

[3] Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, 2017).

[4] Boulanin and Verbruggen (note 3).

*Changing the way the military can field force on the battlefield*

The transformative potential of autonomy derives, first and foremost, from the fact that autonomy can help the military to overcome a number of operational and economic challenges associated with manned operations.[5] According to Scharre, it permits the military to deploy weapons, means and methods of warfare with '1) greater speed, 2) agility, 3) accuracy, 4) persistence, 5) reach, [and] 6) coordination and mass'.

1. *Greater speed*. Autonomy can enable weapon systems to execute the so-called observe, orient, decide, act (OODA) loop much faster than any human ever could, which explains why autonomy is deemed particularly attractive for time-critical missions or tasks such as air defence (detecting and targeting high-velocity projectiles) and air-to-air combat (between combat aircraft). It is also well suited to cyber-defence (discovering and neutralizing a cyberattack) and electronic warfare (analysing and countering new enemy signals).

2. *Agility*. Autonomy can make weapon systems far more agile from a command-and-control perspective and reduce the need to maintain a constant communication link between the system and the military command; it can also allow the military to scale down on the number of human operators and analysts needed to oversee the system and process information.

3. *Accuracy*. Autonomy improves the accuracy of weapon systems, which increases the ability of the military to ensure that the weapon hits only the lawful target or hits the lawful target and causes an acceptable level of collateral damage.

4. *Persistence*. Autonomy improves a weapon system's persistence, meaning that its performance remains unaltered over time. The performance level of a weapon system that is destined for so-called dull, dirty or dangerous missions (3D tasks), such as air defence, long surveillance missions, countermine operations or logistics operations in enemy territory, might deteriorate over time due to the human operator's cognitive and physical limitations, linked to fatigue, boredom, hunger or fear. Autonomy removes these limitations.

5. *Reach*. Autonomy can give weapon systems greater reach. It grants access to operational theatres that were previously inaccessible to remotely controlled unmanned systems or too risky for soldiers or manned systems. Such theatres include anti-access/area denial (A2/AD) bubbles and areas where there are harsh operating environments for humans (and where communication is limited), such as deep water, the Arctic and, potentially, outer space.

6. *Coordination and mass*. Autonomy also provides new opportunities for collaborative operations as it permits weapon systems to operate in large groups, or 'swarms', in a much more coordinated, structured and strategic way than if they were individually controlled by a human operator.

It is important to note that autonomy provides operational benefits across a very broad range of missions, from defensive, intelligence surveillance, reconnaissance and logistics missions to combat missions. Weapons, means and methods of warfare with autonomous capabilities can therefore be very different in terms of purpose and technical characteristics.

*Changing the way people make decisions and take action on the battlefield*

Another remarkable effect of autonomy on the conduct of modern warfare is how it changes the way humans interact with the battlefield, or more specifically how they

[5] Scharre (note 2).

make lethal decisions and act on these decisions. Mindell explains that progress in autonomy is changing human decision making and action in at least the following three ways.[6]

1. *Location of decision and action*. Autonomy makes it possible to physically decouple soldiers from their weapons. They may be far away from the battlefield when they decide to deploy them.
2. *Timing of decision and action*. Autonomy also provides an opportunity for soldiers to modulate over time the results of their decisions. They could for instance program a weapon system to take action minutes, days or even months after its activation.
3. *Nature of decision and action*. Autonomy changes the very nature of the decisions and actions that soldiers are supposed to make. Advances in autonomy have created situations in which soldiers must shift from the active role of a 'controller' to the passive role of 'supervisor'. The skillsets, character traits and training that soldiers need have consequently also evolved.

A key takeaway for the legal analysis is that autonomy transforms the way people interact with weapon systems and make decisions on the battlefield, but does not necessarily eliminate their role. Weapon systems will never be 'fully' autonomous in the sense that freedom of action will always be controlled by humans at some level, and programming will always be the product of human plans and intentions. The issue of human control might, therefore, be an essential element to consider in an Article 36 review of a weapon, means or method of warfare that features some form of autonomous capability.

## II. Autonomy and the Article 36 review process

### Rules and principles of international law applicable to the review of autonomous technologies

As mentioned above, autonomy raises different legal concerns depending on the types of functions and decisions that are being made autonomous. For obvious reasons, the most fundamental legal concerns from the perspective of an Article 36 review arise when autonomy supports the targeting process.[7] However, autonomy raises a different set of issues if it is being used (*a*) to execute the targeting process entirely autonomously, as would be the case for autonomous weapon systems; or (*b*) as a decision aid to help humans execute the targeting process.

#### *Autonomous weapon systems*

The targeting process requires a complex qualitative and quantitative assessment to ensure that an attack is in accordance with the fundamental rules and principles of IHL in the conduct of hostilities. This is also known as the law of targeting and comprises distinction, proportionality and precaution in attack (see box 4.1).[8] A truly autonomous weapon system would have to be capable of following each of these three rules to be considered lawful.

Some weapon systems can already distinguish between 'simple' target types autonomously. However, no existing weapon system has a sufficient level of situational

[6] Mindell, D., *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (Viking: New York, 2015).

[7] The ICRC has identified 4 critical functions in the targeting process: from target acquisition to target tracking, target selection and target engagement (or weapon release). ICRC, *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, Expert Meeting Report (ICRC: Geneva, 2014).

[8] On the rules on distinction see Articles 41, 48, 50 and 52 of Additional Protocol I; on proportionality see Article 51(2) of Additional Protocol I; on precaution see Article 57(2).

awareness to autonomously evaluate military advantage or expected collateral damage against military advantage. In many respects, it is highly debatable whether this will ever be technically possible in the future.[9] Given the current state of technology, it is impossible for a system to make the kind of qualitative assessment necessary to determine whether an attack is proportionate or excessive in relation to the military advantage anticipated.

This does not mean that the use of autonomous weapon systems is unlawful per se. Existing systems that select and engage targets without the direct involvement of a human operator indicate that the operation of autonomous weapon systems could be lawful if humans can (*a*) determine the type of target and undertake the proportionality assessment before the launch of the system; and (*b*) predetermine the maximum amount of acceptable collateral damage linked to specific targets.[10] However, this is currently only possible in a very limited number of circumstances: when the target type is unequivocally military and the operational environment is predictable and unlikely to change quickly, such as in an open environment where there are no civilians or civilian objects. Based on the current level of technology, the use of an autonomous weapon system against human targets or in a complex and dynamic operational context such as an urban area would be unlawful unless human operators maintained direct control or oversight over the weapon system's behaviour.

*Autonomy in decision support*

The use of autonomy in decision aids is a common feature of modern weapon systems. Automatic target recognition systems are typically used to help human operators identify, acquire, track, cue or prioritize targets. The fact that autonomy is used in a supporting role, and that a human is the final arbiter in the targeting process, raises a different set of legal issues for the review process. In this case, the key question is not whether the system can comply with the rules of targeting, but whether the system is reliable enough to ensure that the use of the weapon is in compliance with the requirements of international law. The following questions should therefore be considered during the review.

1. What type of information is the system feeding to the human operator and how reliable is that information?
2. Is the system's behaviour comprehensible to trained operators?
3. What are the technical limitations of the system and how do these affect the use of the weapon by a human operator?

These questions are important because they condition the combatant's ability to use the weapon lawfully. It is not difficult to imagine that a lack of reliability in the recognition software, a poorly designed human–machine interface or a lack of training could lead to a situation in which the combatant using the weapon might end up engaging unlawful targets or causing superfluous injury or unnecessary suffering.[11]

[9] Docherty, B., *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch/International Human Rights Clinic: Washington, DC, 2012); Sharkey, N., 'Towards a principle for the human supervisory control of robot weapons', *Politica & Società*, no. 2 (May–Aug. 2014), pp. 305–24; and Sharkey, N., 'Saying "no!" to lethal autonomous targeting', *Journal of Military Ethics*, vol. 9, no. 4 (2010).

[10] These include the Phalanx Close-in Weapon System (CIWS), the Brimstone anti-tank missile and the Harpy loitering munition. Humans deploy these systems but they can fire at targets automatically on the basis of predetermined parameters. Thurnher, J., 'Means and methods of the future: autonomous systems', eds P. A. L. Ducheine, M. N. Schmitt and F. P. B. Osinga, *Targeting: The Challenges of Modern Warfare* (Asser Press: The Hague, 2016), p. 189; and Schmitt, M., 'Autonomous weapon systems and international humanitarian law: a reply to the critics', *Harvard National Security Journal*, vol. 73, no. 2003 (2012), pp. 19–20.

[11] It has been established that when a weapon automatically acquires and cues a target for a human operator, there is a risk of 'automation bias', whereby 'the human operators come to accept computer generated solutions as correct

**Box 4.1.** Targeting law

The principle of *distinction* requires a determination regarding whether a target is lawful and hence not a civilian or civilian object or a person *hors de combat*. It is accompanied by the principles of *proportionality* and *precaution*. The principle of proportionality prohibits an attack that might be expected to cause incidental loss of life, injury to civilians, damage to civilian objects or a combination of these, which would be excessive in relation to the concrete and direct military advantage anticipated. The principle of precaution stipulates that those who plan or decide on an attack must (*a*) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects, or subject to special protections but are truly military objectives; and (*b*) take all feasible precautions in the choice of means and methods of attack with a view to avoiding or minimizing injury to civilians and damage to civilian objects.

*Source*: Boothby, W., *Weapons and the Law of Armed Conflict* (Oxford University Press: Oxford, 2009).

### *Additional considerations*

A legal review is normally only intended to assess the legality of a weapon, means or method of warfare under the normal and planned circumstances of its use. There is no obligation to review situations where the systems might, for instance, be captured and used by the enemy. In the case of autonomous systems, however, it might make sense to factor into the review the risks of unintended harm in case of a system malfunction or unintended loss of control, caused, for instance, by a cyberattack or a programming error.[12] This concern is not limited to the automation of the targeting process. A systems error or a cyberattack leading to a failure of the automated or autonomous flight control mechanism of a loitering munition could lead to a crash in a populated area and cause incidental injury to civilians or damage to civilian objects.[13]

## Checklist for the review of weapons and means of warfare that build on autonomous capabilities

The key legal questions that a reviewing authority would need to consider in a review of weapons and means of warfare that feature autonomous capabilities can be summarized as follows.

1. With regard to its technical characteristics, capabilities and intended effects in normal and planned conditions of use, can it be established that the weapon system is capable of compliance with international law? On this basis, the reviewing authority should consider (*a*) whether the weapon or its use in normal conditions could cause unnecessary suffering or superfluous injury to combatants, injure or damage lawful targets, civilians or civilian objects indiscriminately, or cause long-term, widespread or severe damage to the natural environment; and (*b*) whether the weapon—if it can select and fire at targets autonomously—would have the capacity to comply with the principles of distinction, proportionality and precaution in attack within the specific context of an operation. If the weapon does not have that capacity, what restrictions can be placed on the use of the weapon? If it does have the capacity to comply, to what

and disregard or don't search for contradictory information'. Automation bias is a well-known phenomenon that can lull the operator into engaging in an unlawful attack by attacking the wrong target. Sharkey, 2014 (note 9). See also Murphy, R. and Burke, J., 'The safe human–robot ratio', eds M. Barnes and F. Jentsch, *Human–Robot Interaction in Future Military Operations* (CRC Press: Boca Raton, FL, 2010); and Parasuraman, R., Molloy, R. and Singh, I. L., 'Performance consequences of automation-induced "complacency"', *International Journal of Aviation Psychology*, vol. 3, no.1 (1993).

[12] Automation and autonomy are enabled at the software level. The system's ability to perform the task it has been assigned—in this case targeting but also other operational tasks such as take-off or landing or flight control—will depend on the quality of its computer code. The more complex the task, the more complex the defining algorithm will have to be. As the code and algorithm increase in complexity, the risk increases of a programming error that could lead to a systems failure or provide adversaries with a vulnerability to exploit during cyber-offensive operations.

[13] States also have a strategic interest in ensuring that if a system fails it will 'fail safe'—so that an enemy will not be able to reuse it for hostile purposes.

extent does automation improve distinction, proportionality and precaution in the application of force compared with existing weapons?

2. If the weapon can select and fire at a target autonomously, in what circumstances might the use of this system (*a*) constitute a violation of the right to life or the right to dignity of the target; (*b*) be considered unacceptable under the principles of humanity and the dictates of public conscience; or (*c*) create a responsibility gap in the case of a violation of international law?

Depending on the response to these questions, the reviewing authority could place restrictions or make recommendations on the use of the system. These could be integrated into the programming of the system, the rules of engagement and training programmes. Such restrictions and recommendations might include the following.

1. Restrictions on the operational environment, such as limiting use to a specific domain, a predefined type of location (e.g. an unpopulated area) and/or a particular time span in which the system must locate and engage the target. These restrictions could be directly programmed into the systems.

2. Recommendations on the human–machine command-and-control relationship, such as requesting constant supervision by a human operator and enabling that operator to abort the mission or override the actions of the system.

*Testing and evaluation*

As is the case with any type of weapon, the reviewing authority should review the empirical evidence from the results of robust and unbiased testing and evaluation of the system under review. However, autonomy introduces new types of requirements. Besides ensuring that the intended effects of the weapon do not cause indiscriminate damage, unnecessary suffering or superfluous injury, or have long-term, widespread and severe impacts on the environment, the testing and evaluation should determine the following.

1. The autonomous system as a whole or relevant autonomous functions perform as anticipated and intended in normal conditions, that is, the system or its functions are capable, effective, reliable and suitable for the tasks assigned.

2. The autonomous system includes a safety and or anti-tamper mechanism that minimizes the possibility or consequences of unintended loss of control due to systems failure or cyberattack.

3. The potential consequences of accidental misuse, loss of control, systems failure or cyberattack are foreseeable.

4. If the weapon can select and fire at a target autonomously, it has been correctly programmed to respect the requirements of international law or to allow the end user to exert adequate and meaningful human control over its actions. This requires that the behaviour of the system is predictable and understandable to trained operators, that it can provide traceable feedback on its status and that an operator can activate and deactivate the system.

5. If the system uses autonomous functions as a decision aid for the targeting process, their interface allows the human operator to execute the mission in compliance with the requirements of international law. Here again, this entails that the behaviour of the system is predictable and understandable to trained operators, that it can provide traceable feedback on its status and that an operator can activate and deactivate the system.

## Outstanding issues

Conducting a legal review of weapons, means and methods of warfare with autonomous capabilities presents a number of challenges. These include complexity and cost, the sustainability of the review when a system has the capacity to learn, and the standard of acceptance.

*Complexity and cost of the testing and evaluation procedure*

Testing and evaluation of weapons and means of warfare with autonomous capabilities become increasingly complicated and costly as the systems and the operating environment grow in complexity. To test and evaluate the performance and reliability of an autonomous system, it is necessary to conduct separate validation and verification procedures on the hardware—principally the sensors—and the software, and then to conduct testing and evaluation of the system as a whole. The more complex the architecture of the system, the harder and more expensive it becomes to do separate tests and evaluations. Assessing the performance of a system can be further complicated if it is to be used in a system-of-systems. In such cases, it is essential to take account of the fact that the performance of the system will be affected by the performance of other systems, such as satellites and ground radar. The complexity of the environment in which the system is intended to operate is another important variable that further complicates and increases the cost of the testing and evaluation process. Evaluation of performance and the level of risk associated with the use of a system must be done in the context of various realistic operational scenarios. The more complex the operating environment, the harder and more expensive it becomes to create a representative and realistic scenario for the purpose of testing and evaluation. As the complexity and cost of testing and evaluation grow, the number of states that have the appropriate expertise and resources to independently fund the required types of testing and evaluation is likely to decrease.

*Sustainability of the review: the challenge of learning systems*

While existing applications for autonomy in weapon and military systems are fairly static in their design, progress in the field of machine learning is expected to lead to the introduction of weapon systems and means of war that are capable of learning online—or during their deployment—in the relatively near future. As noted above, the introduction of a learning capability would make the system capable of automatically re-parameterizing and reprogramming itself. From the perspective of Article 36, this raises a new practical problem: when would this internal change of properties count as a modification requiring a new review?

There is little doubt that a new review would have to be conducted if the change in the programming were such that it would make the system operate in a way not considered by the first review, or if it changed the system's or end user's ability to comply with legal standards. The problem is that this type of assessment is not feasible either practically or financially under existing methods of testing and evaluation. New methodologies for testing and evaluation would therefore have to be developed. Some states, in particular the United States, are currently conducting important research and development (R&D) efforts to resolve this issue. One of the paths that is currently being explored is the development of a method for continuous testing and evaluation. A system would report on its learning and modifications throughout its lifecycle, allowing the end user to understand whether a new legal review might be needed. The creation of such 'self-explainable' or 'self-aware' autonomous systems remains at this stage an R&D objective rather than a reality. Any new weapon or means of

warfare that featured online learning behaviour would not in the current technological circumstances be expected to pass a review.

*Standard of acceptance*

A final, fundamental challenge is the difficulty in determining the level of evidence of a system's reliability that a reviewer would need in order to certify the system as lawful to use. As Backstrom and Henderson explain, 'quantifying reliability is not a "yes" or "no" proposition, nor can it be achieved by a single pass/fail test, but rather it is subject to statistical confidence bounds'.[14] A test might indicate that a weapon with an automated target recognition function recognizes a lawful target 30, 80 or 100 per cent of the time. Determining the data-match acceptance criterion is incontestably a sensitive issue. Should it be 100 per cent? Would it be sufficient if the system could identify the correct target 80 per cent of the time? Determining the standards of acceptance with regard to systems failure is incontrovertibly a sensitive issue. For obvious security reasons, states are reluctant to disclose the standards of acceptance that they apply. At the same time, however, this lack of information generates mistrust in the Article 36 review mechanism, particularly from civil society. It seems to be generally agreed, however, that a determination of an acceptable level of a weapon's reliability can only be made on a case-by-case basis, as it will depend on the intended use, the type of target, the expected effect of the weapon and the environment in which it is used. One variable that it is particularly important to consider is the foreseeable effects of weapon failure. The more dramatic these are, the more conservative the standard of acceptability needs to be.

[14] Backstrom, A. and Henderson, I., 'New capabilities in warfare: an overview of contemporary technological developments and the associated legal and engineering issues in Article 36 weapon reviews', *International Review of the Red Cross*, vol. 94, no. 886 (Summer 2012).

# 5. Reviewing the legality of military human enhancement technologies

A third emerging technology area that is expected to shape the future of warfare is the enhancement of military personnel, known as 'military human enhancement' (MHE). MHE technologies are not stand-alone weapons or means of warfare, but are technologies designed to improve human warfighting capabilities. It is therefore not obvious whether, and if so how, the requirements of Article 36 apply in this area.

## I. A short introduction to military human enhancement technologies

### What is military human enhancement?

*Definition*

'MHE' has been defined as 'the process of endowing an individual with an ability that goes beyond the typical level or statistically normal range of functioning for humans generally (or the personal unenhanced capabilities of a particular individual), where the ability is either integrated into the body or is so closely worn or connected that it confers an advantage similar to an internal or organic enhancement that transforms the person'.[1] It therefore refers to processes that elevate human capabilities beyond normal standards.

Determining what falls within the definition of MHE can be difficult, however, given that most enhancement techniques can also be used as therapy in the process of restoring a human to normal ability.[2] The distinction is particularly difficult to make when the technique is used to improve soldiers' resilience to disease, injury or psychologically traumatic events. Furthermore, there is no way to clearly define a normal healthy state, as this is highly variable and subjective.[3]

### Trends in military human enhancement technologies

*Typology: methods and objectives of military human enhancement*

Military personnel can be enhanced in different ways and for different purposes. Enhancement can be accomplished through the use of pharmaceuticals, vaccines, gene editing, exoskeletons, body implants and a variety of other techniques derived from scientific advances in the fields of nanotechnology, neuroscience and biotechnology, and especially synthetic biotechnology.

Methods of MHE can be roughly divided into the following three categories depending on whether they are destined to primarily affect physical, psychological or cognitive capabilities.[4]

1. *Physical enhancements*. These seek to improve soldiers' speed, strength, endurance or senses.

[1] Harrison Dinniss, H. A. and Kleffner, J. K., 'Soldier 2.0: military human enhancement and international law', *International Law Studies*, vol. 92 (2016), p. 434.

[2] Juengst, E. T., 'Can enhancement be distinguished from prevention in genetic medicine?', *Journal of Medicine and Philosophy*, vol. 22, no. 2 (Apr. 1997), p. 29; Daniels, N., 'Normal functioning and the treatment-enhancement distinction', *Cambridge Quarterly of Healthcare Ethics*, vol. 9, no. 3 (2000), p. 313; and Lin, P. and Allhoff, F., 'Untangling the debate: the ethics of human enhancement', *NanoEthics*, vol. 2, no. 3 (2008), pp. 253–55.

[3] Bostrom, N. and Roache, R., 'Ethical issues in human enhancement', eds J. Ryberg, T. S. Petersen and C. Wolf, *New Waves in Applied Ethics* (Palgrave Macmillan: New York, 2007), p. 137.

[4] For a detailed overview of different types of research programmes see Mehlman, M., Lin, P. and Abney, K., *Enhanced Warfighters: Risk, Ethics, and Policy* (California Polytechnic State University: San Luis Obispo, CA, 2013), pp. 22–27.

2. *Psychological enhancements*. These seek to alter soldiers' emotions or ability to deal with various types of emotion.[5]

3. *Cognitive enhancement*. This aims to alter cognitive capabilities, such as intelligence and decision-making ability, awareness, attention, memory, planning and the learning of new skills. Non-invasive brain–computer interfaces, for example, can be used to improve cognitive skills in areas such as learning languages.

Enhancements are not always for the sole benefit of soldiers. Methods of cognitive enhancement, such as brain implants, might for instance be intended to allow a commander to directly monitor, and exercise control over, the mental state and behaviour of his or her troops.

*State of play*

The field of MHE is still at its very early stages, although some techniques, such as the use of pharmaceuticals like modafinil and propranolol to help soldiers with stress, have been used for decades.[6] Many techniques of enhancement have not gone beyond the research and experimentation phase and are not expected to be used by soldiers in the coming decade. (This is particularly true for techniques that involve gene engineering and neurological enhancement through invasive techniques, such as brain–computer interfaces.[7]) Notable exceptions are certain types of exoskeleton and artificial robotic limbs, which are technologies that have reached the stage of marketable application.[8]

## II. Military human enhancement and Article 36

### Applicability of Article 36 to military human enhancement: what is to be reviewed?

In what circumstances would the introduction of new MHE technologies or methods trigger an Article 36 review? It is widely agreed in the literature that enhanced soldiers could not be viewed as weapons, since the concept only refers to objects.[9] Treating combatants as weapons would contradict the good-faith interpretation and ordinary use of the term. For Harrison Dinniss and Kleffner, it would also raise a range of complicated legal issues and have dangerous implications for how soldiers might be treated, as people are normally afforded substantially more protection than objects.[10]

The fact that soldiers are not reviewable does not mean that Article 36 has no part to play. For Harrison Dinniss and Kleffner, the enhancement itself is reviewable as one of the following.

[5] Ford, K. and Glymour, C., 'The enhanced warfighter', *Bulletin of the Atomic Scientists*, vol. 70, no. 1 (Jan. 2014), p. 49; and Mehlman, M. J. and Li, T. Y., 'Ethical, legal, social, and policy issues in the use of genomic technology by the US Military', *Journal of Law and the Biosciences*, vol. 1, no. 3 (Sep. 2014), p. 249.

[6] Modafinil increases brain function so that soldiers are better able to operate under stress. Propranolol decreases particular brain functions so that soldiers can better deal with the stress of battle.

[7] Miranda, R. A. et al., 'DARPA-funded efforts in the development of novel brain–computer interface technologies', *Journal of Neuroscience Methods*, no. 244 (Apr. 2015).

[8] Freedberg, S., 'Lockheed exoskeleton gives troops a leg up, literally', Breaking Defense, 17 May 2017; and Herr, A. and Cheney-Peters, S., *Between Iron Man and Aqua Man Exosuit: Opportunities in Maritime Operations*, 20YY Series (Jan. 2015).

[9] One notable exception in the literature is Mehlman, Lin and Abney (note 4), pp. 22–27. They argue that enhanced warfighters could be considered as weapons for several reasons. First, they state that other organisms used by militaries, such as dogs, sea lions and pigeons, could be plausibly subject to Article 36 reviews. Second, physical implants and replaced body parts could turn soldiers into cyborgs, on a spectrum between humans and machines. If autonomy in weapon systems is subject to review, it makes sense to them that soldiers with cybernetic enhancements would be too, as it would be difficult to identify the point where a human becomes a robot or a weapon. Third, genetically engineering humans could potentially lead to humans so enhanced that they no longer resembled human beings, for instance if they had fangs or multiple arms.

[10] Harrison Dinniss and Kleffner (note 1), pp. 436–39.

1. A weapon—if the enhancement is specifically designed to cause injury or death to enemy personnel or cause damage or destruction to an object. This would be the case for an artificial limb with an inbuilt gun.

2. A means of warfare—if the technology is part of a weapon system and specifically used for the purpose of an attack. This would be the case for a brain–computer interface that allowed an operator to control a weapon with his or her mind or in the case of an eye lens that helped a soldier to identify targets.

3. A method of warfare—if, and when, the enhancement's 'use constitutes an integral part of offensive activities at the strategic and tactical levels'.[11] A biochemical enhancement using drugs such as modafinil or propranolol would not be reviewable as a means of warfare as such an enhancement is not specifically designed to cause injury or death to enemy personnel or damage or destruction to an object. An enhancement could, however, be reviewed as a method of warfare if it is established that the purpose of the enhancement is to make the rule relating to distinction or proportionality impossible to apply.

In short, the determination of whether an enhancement technique is reviewable should be made a priori because it will depend on the end use.[12]

## The rules and principles of international and national law applicable to the review of enhancement technologies

### *Prohibition on superfluous injury or unnecessary suffering*

None of the current or realistically foreseeable enhancement technologies are prohibited by existing weapon treaty law. Some of the cardinal principles of IHL would be particularly important to consider in the review of certain types of enhancement, starting with the general prohibition on weapons, means and methods of warfare capable of causing superfluous injury or unnecessary suffering.

The prohibition on weapons, means and methods of warfare of a nature that might cause superfluous injury or unnecessary suffering is doubly relevant as it protects the enhanced soldier from the effects of both the enhancement and the weapons, means and methods of warfare designed specifically to counter MHE technologies.[13] In that regard, the reviewing authority should take account of the following questions.

1. Is the enhancement process itself of a nature that might cause superfluous injury or unnecessary suffering? Would its introduction, use or removal be particularly painful?

2. If the effects of the enhancement are irreversible—for example in the case of gene editing or the physical modification of body parts—how would this affect the soldier's return to civilian life?[14] Would the enhancement cause permanent psychological torment, disability or disfigurement?[15]

3. Would enemy actions targeted at countering the enhancement technology be likely to cause superfluous injury or unnecessary suffering?

[11] Harrison Dinniss and Kleffner (note 1), pp. 436–39.

[12] This is also the interpretation of Boothby. He states that, as a general rule, enhancement technologies would not be subject to review, unless they constitute a new method of warfare, or the enhancement technology is part of a weapon system or materially changes the manner in which hostilities are prosecuted. This would for instance be the case with a human–machine interface. Boothby, W. H., *Weapons and the Law of Armed Conflict*, second edn (Oxford University Press: Oxford, 2016), pp. 125–26.

[13] Krishnan, A., *Military Neuroscience and the Coming Age of Neurowarfare* (Routledge: London, 2017), p. 224.

[14] Krishnan (note 13), p. 224.

[15] Faunce, T. A. and Nasu, H., 'Nanotechnology and the international law of weaponry: towards international regulation of nano-weapons', *Journal of Law, Information and Science*, vol. 20 (2010), pp. 40–41.

*Prohibition on weapons, means and methods of warfare capable of causing widespread and long-term damage to the natural environment*

The general prohibition on weapons, means and methods of warfare of a nature likely to cause widespread, long-term and severe damage to the natural environment might also be relevant to consider with regard to some methods of enhancement, particularly those that build on advances in nanotechnology and, to a lesser extent, synthetic biotechnology and genomics.[16] Some preliminary studies have reportedly indicated that these technologies might have long-lasting damaging effects.[17] Customary IHL states that a lack of scientific certainty does not absolve a party from taking all feasible precautions to minimize incidental damage to the environment. With regard to the conduct of Article 36 reviews, this means that the reviewing authority is under an obligation to fund or ask for adequate scientific studies to be undertaken to provide evidence of the possible effects of these technologies on the natural environment.

*Expected effect on a soldier's ability to comply with international humanitarian law*

The rules of IHL on the conduct of hostilities are also important as some enhancements, most notably psychological and cognitive enhancements, could affect a soldier's ability to comply with IHL rules, which could, by extension, have an impact on the possibility of assigning criminal responsibility. A soldier may only be found guilty of a violation of IHL if he or she passes the *mens rea* test, meaning that he or she 'possesses a culpable state of mind'.[18] Whether enhanced soldiers would have the ability to form the requisite *mens rea* might be difficult to establish in cases where the enhancement directly affects a soldier's sensory perception, morality, rationality or understanding of social norms.[19]

*International human rights law and enhanced personnel*

IHRL also forms part of the corpus of international law that needs to be considered in the conduct of an Article 36 review. Enhanced soldiers, by virtue of being alive in the world, possess human rights. Their rights may be restricted by the state, particularly during an armed conflict, but not completely abrogated. A number of rights and freedoms could be affected by the use of human enhancement. The following rights and freedoms would need to be considered in a review.

1. *The right to bodily integrity and freedom from torture, inhuman and degrading treatment*. This right may give the soldier the right to refuse certain types of physical enhancement. It may also place some limit on the use of enhancement to cause deprivation for the soldier, such as sleep deprivation.
2. *The right to privacy and freedom of thought and expression*. This right would be affected by human enhancement technologies such as brain–computer interfaces or optical prosthetics, which manipulate information between the outside world and the brain or inhibit the ability to make free choices. The Article 36 review might have to place some restrictions on the effects these have on the enhanced personnel when no longer on duty.

[16] Enhancing an individual does not count as modifying the environment. The rule is limited to 'the dynamics, composition or structure of the Earth, including its biota, lithosphere, hydrosphere and atmosphere, or of outer space'. Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques, signed 18 May 1977, entered into force 5 Oct. 1978.

[17] Coenen, C. et al., 'Human enhancement study', European Parliament Science and Technology Options Assessment, IP/A/STOA/FWC/2005-28/SC35, 41 & 45, May 2009.

[18] Leveringhaus, A., 'Assigning responsibility in enhanced warfare', eds J. Galliott and M. Lotz, *Super Soldiers: The Ethical, Legal and Social Implications* (Ashgate: Farnham, 2015), p. 148.

[19] White, S. E., 'Brave new world: neurowarfare and the limits of international humanitarian law', *Cornell International Law Journal*, vol. 41 (2008), pp. 186–204; and Harrison Dinniss and Kleffner (note 1), pp. 476–82.

3. *Fair trial rights*. Psychological and cognitive enhancements that affect a soldier's memory could threaten that soldier's right to a fair trial, and would create doubt as to his or her reliability as a witness.

4. *The right to life*. This right could be infringed by experimental enhancement technologies that could cause accidental death.

*Regulations on research on human subjects*

Medical research on human subjects is limited by the 1947 Nuremberg Code, the 1948 Declaration of Geneva and the 1964 Declaration of Helsinki.[20] They state, among other things, that soldiers have the right to informed consent and should receive information about the potential risks and benefits of the enhancements, and that the risks must be in proportion to the expected benefits.[21] These would be relevant when assessing the legality of R&D projects.

*National regulations*

There are many relevant domestic laws on the treatment of soldiers and conducting medical research that might also come into play in the review process. These cannot be discussed in detail here. It should be noted that they vary considerably from country to country. For instance, using amphetamines to increase the performance of warfighters is legal in the USA and the UK but not in Denmark or Germany.[22]

## Checklist for the review of the legality of military human enhancement technologies

In sum, the key questions that need to be considered when determining whether, and if so how, an enhancement should be reviewed by an Article 36 procedure are as follows.

1. Is it enhancement or therapy? To what extent does the enhancement improve abilities beyond the normal?

2. Can the enhancement be considered to be (*a*) a weapon, means or method of warfare; or (*b*) specifically designed, or part of a system designed, to cause injury or death to enemy personnel or damage or destruction to an object?

3. Is the enhancement materially changing the manner in which hostilities are prosecuted?

4. What are the physical, biological, physiological, psychological and social effects on the operators, bystanders and targets?

5. Is the enhancement of a nature to cause superfluous injury or unnecessary suffering?

6. Could the methods of enhancement have different effects on different people? Would they affect a woman's body differently?

[20] The Nuremberg Code was developed during trials of war criminals after World War II. Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law no. 10, vol. 2, (US Government Printing Office: Washington, DC, 1949) pp. 181–82. Declaration of Geneva, adopted by the 2nd General Assembly of the World Medical Association, Geneva, Sep. 1948. Declaration of Helsinki, Ethical Principles for Medical Research Involving Human Subjects, adopted by the 18th General Assembly of the World Medical Association, Helsinki, June 1964.

[21] Note that there is substantial debate about the extent to which soldiers are capable of providing informed consent within the command structure of the military. The full detail of this debate is beyond the scope of this report. For more information see Parasidis, E., 'Justice and beneficence in military medicine and research', *Ohio State Law Journal*, vol. 73, no. 4 (2012); and Annas, C. L. and Annas, G. J., 'Enhancing the fighting force: medical research on American soldiers', *Journal of Contemporary Health Law and Policy*, vol. 25, no. 2 (2008).

[22] Future Analysis Branch of the German Armed Forces, *Human Enhancement: A New Challenge for the Armed Forces?* (Bundeswehr: Berlin, 2013), p. 2; and Nielsen, J. N., 'Danish perspective: Commentary on "Recommendations for the ethical use of pharmacological fatigue countermeasures in the US military"', *Aviation, Space, and Environmental Medicine*, vol. 78, no. 1 (May 2007), p. 135.

7. What would be the physical and psychological impact of the potential removal of the enhancement?

8. Would the enhancement be of a nature likely to cause widespread, long-term and severe damage to the environment?

9. Would the enhancement affect a soldier's ability to comply with the rules on the conduct of hostilities? If so, how might this affect his or her ability to pass the *mens rea* test?

10. Would recourse to the method of enhancement infringe the soldier's human rights? If so, what can be done to prevent such infringements?

Depending on the responses to these questions, the reviewing authority could place restrictions or make recommendations that would be integrated into the R&D phase, the rules of engagement and training programmes. Such restrictions and recommendations might include the following.

1. Recommendations on the research and experimentation on human soldiers necessary for the development of the capability, including pre-testing safeguards such as a review of risks and benefits by an ethical committee or measures to ensure informed consent, and mechanisms for adequate outcome monitoring, including long-term monitoring.

2. A requirement that the effect of the enhancement is not irreversible or that it does not undermine the enhanced soldier's reintegration into civilian life.

# 6. Conclusions: Article 36 and technological change

## I. Cross-cutting challenges

Technologies in the areas of cyberwarfare, artificial intelligence and robotics, and human enhancement are at various levels of maturity, from mature to still emerging and experimental. It is beyond dispute, however, that these three technology areas will have a dramatic impact on the future of warfare as they all have the potential to fundamentally change the way force is applied and critical decisions are made on the battlefield. This report has explored the potential and actual impacts of ongoing developments in these fields on the legal review of new weapons, means and methods of warfare. It has established that despite their technical and operational differences, the military applications derived from these technology areas raise similar challenges as far as the conduct of Article 36 reviews is concerned.

### A need to examine old legal concepts anew

Such technologies require a re-examination of old legal concepts. This is particularly obvious in the case of cyberwarfare technologies, which invite a rethink of what have been unequivocal concepts in international law, such as 'weapon', 'attack' or 'armed conflict'. Advances in artificial intelligence, robotics and human enhancement also raise a number of conceptual, if not philosophical, questions about how to apply concepts designed for human agency to machines, and vice versa. How should an autonomous weapon system relate to the key principles of the law on targeting: distinction, proportionality and precaution in attack? In the case of human enhancement technologies, when does an augmented soldier cease to be human and become a mere weapon for the purpose of the review?

### Legal reviewers need increasingly broad technical expertise

The legal experts conducting the Article 36 review must now have acquired expertise on a much broader range of technologies. They must have a good grasp of computer science, robotics, biotechnology and neuroscience. They may not need to be experts but in order to do their job properly, they do need sufficient understanding of the underlying technologies in the weapons, means and methods of warfare they are to review. This includes being able to help technical personnel translate legal requirements into engineering decisions, and understanding the results of tests and evaluations.

### New methods of testing and evaluation are needed

New methodologies are required for the assessment of performance and the risks posed. Testing and evaluation in the area of traditional kinetic weapons are not necessarily easy but the requirements are straightforward: to provide evidence of reliability and the effects on the human body of specific types of material. Testing of cyberweapons, autonomous weapons or methods of human enhancement faces new and in some cases very challenging requirements. In the case of cyberweapons, for instance, it might be difficult to model the environment in which the weapon is designed to be employed. In the case of autonomous weapon systems, existing methods of testing and evaluation do not yet allow an assessment of the performance and reliability of complex learning systems. In the realm of human enhancement, the limitations are

of an ethical nature. States need to have standards on what is permissible in terms of experimentation on human subjects.

## II. Recommendations

There are, however, a number of mechanisms for identifying concrete and viable solutions to the above-mentioned issues. These include (*a*) building on existing elements of best practice; (*b*) enhancing transparency and cooperation in the area of Article 36 reviews; and (*c*) supporting targeted research.

### Build on existing elements of best practice for the conduct of Article 36 reviews

Some general elements of best practice were identified in the ICRC guide to weapon reviews.[1] The following guidelines are aimed at further addressing the increasingly technical and practical complexities of the review process.

1. *Start early*. The review process should start as early as possible, even as early as during the study for a new weapon project. It should preferably be incorporated into the procurement process at a key decision point, and continue until the weapon has been funded and is being used. It is important to keep a written record of the review so that possible restrictions or recommendations are not ignored.

2. *Multidisciplinary approach*. A legal review is a multidisciplinary process, either formally or informally, with input from various fields of expertise, such as legal, medical, operational or technical. Operational and technical expertise can contribute to an understanding of what the item under review is supposed to do and whether it does so in accordance with the set criteria.

3. *Training*. Military lawyers involved in the review process should receive some technical training, which would improve their understanding of the trends in technology development. It is also important to inform engineers and systems developers about the requirements of international law so they can factor these into the design of weapons or means of warfare.

4. *Testing and evaluation*. Conducting tests and evaluations to assess the possible risks associated with use is a crucial aspect of the review process. These tests can be done in cooperation with the manufacturer and the procurement agency. Increased interaction between lawyers, systems developers, technical experts and end users throughout the review process will be instrumental in enhancing all parties' understanding of how testing and evaluation procedures should be developed and interpreted. When systems are complex and expensive, relying on computer simulations is a useful solution to reducing the cost of the procedure.

### Strengthen transparency and cooperation in the area of Article 36 reviews

For obvious strategic reasons, states may be reluctant to share with each other or with the general public detailed information about individual reviews. They might not want to reveal what is or what is not in their toolbox. They do, however, have a lot to gain from supporting greater transparency on Article 36 review procedures and cooperation on dealing with the challenges posed by emerging technologies.

Increased transparency on Article 36 review procedures could become a virtuous circle in at least the following three ways.

[1] ICRC, *A Guide to the Legal Review of Weapons, Means and Methods of Warfare* (ICRC: Geneva, 2006).

1. *Demonstrating compliance*. First, it would allow states that conduct Article 36 reviews to publicly demonstrate their commitment to legal compliance.
2. *Supporting more widespread and robust compliance*. Second, it would be of assistance to states that are seeking to set up and improve their weapon review mechanisms, and thereby create the conditions for more widespread and robust compliance
3. *Strengthening confidence building in the area of weapon reviews*. Finally, it could facilitate the identification of elements of best practice and interpretative points of guidance for the implementation of legal reviews, which would strengthen international confidence in such mechanisms.

Cooperation is an effective way to address some of the outstanding conceptual and technical issues raised by emerging technologies. Dialogues, expert meetings and conferences can allow generic issues to be debated and addressed in a manner that does not threaten the national security of any state. Three separate cooperation tracks could be explored.

1. *Track 1*. Bilateral dialogue and closed meetings of experts, such as the Weapon Review Forum organized by the Development Concepts and Doctrine Centre of the British Ministry of Defence, allow practitioners to exchange views and experience on highly practical and sensitive issues. Testing and evaluation are areas where states have a vested interest in cooperation. International cooperation could rationalize costs.
2. *Track 1.5*. Workshops and conferences involving Article 36 review practitioners and relevant experts from civil society, like the one SIPRI convened to prepare this report, are useful for debating generic issues and translating the product of academic discourse and research into concrete legal and technical advice for review practitioners.
3. *Track 2*. Workshops and conferences involving legal and technical experts from academia or research institutions are essential for investigating deeper legal and technical issues and generating innovative responses to the challenges posed by emerging technologies to the conduct of legal reviews and compliance with international law more generally.

**Support targeted research**

Finally, the above-mentioned challenges cannot be overcome without relevant and rigorous research. A number of general outstanding issues deserve the particular attention of scholars and practitioners. These include but are not limited to the following.

1. *Cyberwarfare technologies*. How to apply the concepts of distinction, proportionality and precaution in an environment that is primarily characterized by dual-use technology.
2. *Artificial intelligence and robotics*. How to verify the predictability of autonomous weapon systems' compliance with international law.
3. *Human enhancement*. At what point does an enhanced soldier cease to be a human being and become a mere weapon for the purpose of an Article 36 review?

**STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE**
Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org