Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

28 March 2019

English only

Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems
Geneva, 25–29 March 2019 and 20–21 August 2019
Agenda item 5
Focus of work of the Group of Governmental Experts in 2019

Implementing International Humanitarian Law in the Use of Autonomy in Weapon Systems

Submitted by the United States of America

I. Introduction and Overview

1. This U.S. working paper addresses the implementation of international humanitarian law (IHL) in the use of autonomy in weapon systems. The paper builds on U.S. working papers submitted in 2017 and 2018.[1] Specifically, this working paper discusses: IHL requirements and autonomous functions in weapon systems; steps that States can take to help implement IHL requirements; and the potential for emerging technologies in the area of lethal autonomous weapon systems (LAWS) to strengthen implementation of IHL.

2. In summary, this paper offers the following main conclusions:

(a) Existing IHL, including the requirements of distinction, proportionality, and precaution, provides a comprehensive framework to govern the use of autonomy in weapon systems.

(b) Internal procedures for review and testing, including the legal review of weapons, are essential to implementing IHL requirements.

(c) Emerging technologies in the area of LAWS could strengthen the implementation of IHL, by, *inter alia*, reducing the risk of civilian casualties, facilitating the investigation or reporting of incidents involving potential violations, enhancing the ability to implement corrective actions, and automatically generating information on unexploded ordnance.

[1] U.S. working papers to the CCW GGE include: *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, Aug. 28, 2018, CCW/GGE.2/2018/WP.4; *Humanitarian benefits of emerging technologies in the area of lethal autonomous weapon systems*, March 28, 2018 CCW/GGE.1/2018/WP.4; *Autonomy in Weapon Systems*, Nov. 10, 2017, CCW/GGE.1/2017/WP.6; *Characteristics of Lethal Autonomous Weapons Systems*, Nov. 10, 2017, CCW/GGE.1/2017/WP.7.

II. IHL Requirements and Autonomous Functions in Weapon Systems

3. Although a wide range of IHL requirements could be implicated by the use of autonomous weapon systems,[2] this paper seeks to focus on IHL issues distinctly presented by the use of autonomous functions in weapon systems to select and engage targets.[3] The following IHL requirements are of particular relevance:

(a) Distinction. "Combatants may make military objectives the object of attack, but may not direct attacks against civilians, civilian objects, or other protected persons and objects."[4]

(b) Proportionality. "Combatants must refrain from attacks in which the expected loss of life or injury to civilians, and damage to civilian objects incidental to the attack, would be excessive in relation to the concrete and direct military advantage expected to be gained."[5]

(c) Precaution. "Combatants must take feasible precautions in planning and conducting attacks to reduce the risk of harm to civilians and other persons and objects protected from being made the object of attack."[6]

4. In considering these requirements, it may be helpful first to recall:

(a) These requirements impose duties on combatants (and, more broadly, on parties to conflict), and do not impose duties on machines.

(b) These requirements address "attacks," rather than the firing or activation of weapon systems as such. For example, the single firing of a weapon system might only be one part of an "attack," and the mere activation of a weapon system might not constitute an "attack" at all.

(c) These requirements are implemented in military operations through responsible commands, and not every duty will be implemented by every individual within the command. For example, the decision about whether a particular precaution is feasible might be made by a commander at a particular level of command with the authority to direct the resources necessary to take that precaution, or individual units within the command might be tasked with carrying out a precaution, such as delivering a warning.

(d) Although the implementation of these requirements through the chain of command necessarily will entail different duties being fulfilled by different individuals within the command, all combatants, when prosecuting attacks against military objectives, must exercise due regard to reduce the risk of incidental harm to the civilian population and other persons and objects that may not be made the object of attack. This standard of due regard must be assessed based on the general practice of States and common standards of the military profession in conducting operations.

5. As a threshold matter, how IHL requirements are implicated could depend on how the commander or operator is using the autonomous functions in the weapon system. Based on past practice in using autonomy in weapon systems, we can envision at least three general scenarios for the use of autonomous functions in weapon systems.[7]

[2] The definition of "autonomous weapon system" provided in U.S. Department of Defense Directive 3000.09, Autonomy in Weapon Systems, is one that "once activated can select and engage targets without further intervention by a human operator."

[3] For example, this paper is not intended to focus on the prohibition on killing or wounding by resort to perfidy or the requirements of the CCW Amended Protocol II, although it is possible that the use of autonomous weapon systems could present such issues depending on the nature of the systems and intended use of the weapon systems.

[4] U.S. Department of Defense (DoD) Law of War Manual 5.4.2.

[5] DoD Law of War Manual 5.4.2.

[6] DoD Law of War Manual 5.4.2.

[7] Because emerging technologies continue to be developed; there could be novel uses of emerging technologies that would not be reflected in these scenarios.

(a) First, a weapon system's autonomous function could be used to effectuate more accurately and reliably a commander or operator's intent to strike a specific target or a specific target group. For example, the operator identifies an enemy surface-to-air missile system and fires a missile at it. Rather than only being guided by the operator's aiming of the missile at the target, the missile also has sensors and computers that provide it the capability to recognize enemy surface-to-air missile systems (*e.g.*, through detection of electromagnetic emissions of the enemy surface-to-air missile system and comparison with an onboard "library" of such emission "signatures"), and, after being fired, the missile automatically identifies, acquires, and guides itself to the target that the operator intended to strike.

(b) Second, a weapon system's autonomous functions could inform a commander or operator's decision-making about what targets he or she intends to strike. For example, the computers and sensors on the weapon system could generate an assessment of a potential target that the operator would consider along with other relevant information (*e.g.*, the operational context in which the weapon is deployed) in deciding whether to engage a target. These types of computers and sensors also could, in principle, be distinct from the weapon system used to engage a target. For example, counter-battery radar systems are used to identify the location from which incoming rockets, artillery, and mortars were launched, which is used to direct counter-battery fire by an artillery system.

(c) Third, a weapon system's autonomous function could be used by a commander or operator to select and engage specific targets that the commander or operator did not know of when he or she activated the weapon system. For example, a commander might assess that there is a general risk of enemy missile or rocket attacks against a given location or against a given unit or platform, but the commander might not know of a specific incoming missile or rocket attack. In order to protect that location, unit, or platform, the commander might direct the activation of a weapon system, such as an active protection system, that would select and engage incoming projectiles automatically if such an attack occurs. Similarly, a commander might assess a risk that enemy tanks will attack and deploy a field of anti-tank mines at a strategic location, such as at a mountain pass, to counter that potential threat.

6. In the first scenario described in paragraph 5.a., the analysis of IHL requirements proceeds almost identically as in a case of use of a "dumb" weapon without the autonomous function because the weapon is being used no differently. The addition of autonomous functions, however, is intended to enhance the effectiveness of the weapon system by making it more accurate and precise in striking military objectives.

(a) The addition of autonomous functions would render the use of the missile illegal if the autonomous functions, contrary to the intention of making the weapon more precise and accurate, actually made the missile inherently indiscriminate, *i.e.*, incapable of being used consistent with the principles of distinction and proportionality. On the other hand, it would be reasonable to rely on the autonomous function in the missile to identify, acquire, and guide itself to the target or target group, to the degree that the autonomous function performed accurately and consistently in selecting and engaging the correct targets. For example, if testing indicated that addition of the autonomous function to an already lawful missile system served only to improve its accuracy in striking military objectives, then it clearly would be appropriate to rely on the autonomous functions in the weapon system because IHL does not prohibit increasing the precision of weapons.

(b) This first scenario illustrates that if weapons systems with autonomy in targeting functions are used in the same way as weapon systems lacking such capabilities (*i.e.*, to strike a specific target while complying with IHL requirements), they do not seem to present new issues of IHL compliance. Moreover, when such a reliable autonomous function is available, the use of such a weapon could also be deemed a feasible precaution to reduce the risk of civilian casualties.

7. The second scenario, described in paragraph 5.b., presents a general issue of when it is permissible or appropriate to rely on autonomous functions to aid in decision-making in armed conflict.

(a) Armed conflict, and combat operations in particular, takes place in a difficult decision-making environment – often referred to as the "fog of war." The facts may be difficult to discern due to the efforts by the adversary to deceive as well as the stress and chaos accompanying combat operations, including the constant threat of attack by the adversary. Recognizing that information during armed conflict may be imperfect or lacking, commanders and other decision-makers must make a good faith assessment of the information that is available to them at the time when conducting military operations.[8]

(b) An autonomous function that was intended to provide probative information that increased the accuracy of decision-making, such as the counter-battery radar system, would generally be a permissible tool for the commander or weapon system operator to consider as one of the available sources of information in making decisions.

(c) More specifically, whether it were permissible under IHL to rely on an automated assessment to consider a target to be a military objective would depend on whether such reliance was consistent with the exercise of due regard to reduce the risk of incidental harm to the civilian population and other persons and objects that may not be made the object of attack. This would depend, *inter alia*, on: i) an understanding of how accurately and consistently the machine performs in not mischaracterizing civilian objects as military objectives; ii) the operator giving the automated assessment appropriate weight relative to other information that would be probative of whether the target, in fact, was a military objective; and iii) the urgency to make a decision. For example, if the automated assessment has a very low rate of "false positives" (even relative to the "false positive" rate for a person making such an assessment), the operational context corroborated the automated assessment (*e.g.*, intelligence reporting indicated the possibility of the threat identified by the system), and the context involved combat operations, then it would seem to be reasonable to rely on the assessment to conclude that the target was a military objective and, provided other IHL requirements were met, to strike the target. On the other hand, if the system performed with a significant rate of "false positives," there was no need for a rapid decision, and the contextual factors contradicted the automated assessment, then it would not seem to be reasonable for a person to rely on the automated assessment to conduct a strike immediately, rather than first seeking further information.

8. The use of autonomous functions in a weapon system in the third scenario, described in paragraph 5.c., to select and engage targets may also be consistent with the requirements of distinction, proportionality, and precaution. That is, although the commander or operator would not expressly intend to strike a specific target or target group when activating the weapon system, it may be possible that the weapon system nonetheless could be operated consistent with these IHL requirements. By way of comparison, anti-tank mines similarly may be used consistent with IHL without an express intention at the time of emplacing or activating the mines that the mines detonate against a specific tank.

(a) The commander or operator could act consistently with the principle of distinction by having the intention of making potential military objectives (*e.g.*, the potential incoming projectiles in the active protection system mentioned above) the object of attack and provided that the autonomous functions in the weapon system perform with sufficient reliability (*e.g.*, consistently selecting and engaging the incoming projectiles) to ensure that force can be directed against the potential military objectives. Alternatively, if the commander or operator identified an area that constituted a military objective, the weapon system could be deployed against that area, for example, to divert enemy forces from that area.

(b) The commander or operator could act consistently with the principle of proportionality by assessing that the risk of civilian casualties from the activation of the weapon would not be excessive in relation to the military advantage expected to be gained. An assessment of the risk of civilian casualties could be informed by a variety of factors, including any precautions taken to reduce that risk. For example, if the weapon's autonomous function performed accurately and reliably to fire only against incoming projectiles and used rounds that self-destructed in flight to reduce the potential for the

[8] DoD Law of War Manual 5.3 (June 2015, Updated December 2016).

round to cause harm if it missed the target, then the commander could be able to assess the risk of civilian casualties to be minimal and not excessive compared to the military advantage expected to be gained from activating the system. Warnings to potential civilian air traffic and monitoring the operation of the weapon system could also be important precautions to take that would reduce the risk of civilian casualties and ensure that the activation of the system would not be excessive.

(c) Even if the risk of civilian casualties was not expected to be excessive in relation to the military advantage expected to be gained, it would be important to take further feasible precautions. For example, warnings, monitoring, and self-destruct, self-deactivation, or self-neutralization mechanisms are all precautions that have been usefully employed to reduce the risk of civilian casualties in relation to the use of land or naval mines. These precautions also should be considered in relation to new types of autonomous weapon systems that would be employed in a similar fashion. Moreover, emerging technologies might afford even more capabilities to reduce the risk of civilian casualties. For example, a new type of system that could autonomously target enemy forces entering an area could be more discriminate than existing lawful means of securing that military objective, such as emplacing a marked and monitored minefield or providing interdiction fires through artillery. In such a case, use of the autonomous system might even be regarded as an additional precaution that should be taken, consistent with IHL.

III. Practical Measures to Help Implement IHL Requirements

9. Considering these scenarios helps highlight the importance of practical measures to implement IHL requirements in respect of autonomous functions in weapon systems. These practical measures include: 1) rigorous testing to assess system performance and reliability; 2) establishing doctrine, training, and procedures to ensure that weapons are used in accordance with how they have been designed, tested, and reviewed; and 3) the legal review of weapons prior to their use.

10. Rigorous testing to assess weapon system performance and reliability supports compliance with IHL. For example, in the first scenario described in paragraph 5.a., if the autonomous function performs erratically, randomly engaging civilian objects rather than the intended military targets, then the weapon would appear to be prohibited as inherently indiscriminate. Similarly, in the second scenario described in paragraph 5.b., the degree to which the autonomous function misclassifies civilian objects as military objectives would be a significant factor in whether it would be reasonable to rely on the assessment as the basis for conducting a strike. States have every incentive to develop reliable systems; rigorous testing of systems reflects a strong convergence of military and humanitarian interests.

11. Establishing doctrine, training, and procedures for the weapon system is another important mechanism that helps ensure that the weapon is used consistent with IHL. What IHL considerations are salient could depend on how the weapon is to be used. If a weapon system was developed with a particular concept of employment in mind (*e.g.*, scenario 1 described in paragraph 5.a.), it might create unanticipated problems if the weapon were used in ways not contemplated by those who developed and tested the weapon system (*e.g.*, scenario 3 described in paragraph 5.c.).

12. The legal review of weapons prior to their use enables the State developing or acquiring the weapons to consider relevant IHL issues, including precautions to reduce the risk of civilian casualties. The legal review of the weapon also affords an opportunity to ensure that designers and developers of the weapon system and others tasked with ensuring the reliability of the weapon system have applied their expertise. Similarly, the legal review of the weapon also provides a mechanism for reviewing doctrine, training, and procedures for the weapon system and considering additional doctrine, training, and procedures that would help ensure the weapon is used consistent with IHL.

IV. Potential for Emerging Technologies in the Area of LAWS to Strengthen Implementation of IHL Requirements

13. One of the guiding principles for the GGE's work is that "[c]onsideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations."[9] Thus, it is important to consider how emerging technologies in the area of LAWS can strengthen implementation of IHL requirements.

14. As the United States has noted in its working paper on humanitarian benefits of emerging technologies in the area of lethal autonomous weapon systems submitted at the August 2018 GGE meeting, new advancements in autonomy in weapon systems hold great promise for strengthening the implementation of IHL. For example, as also discussed in this paper, autonomous functions could be used to make weapons more precise and increase the accuracy of human decision-making in stressful and time-critical situations.

15. In addition, emerging technologies in the area of LAWS could have benefits that extend beyond simply reducing the risk of civilian casualties in military operations.

(a) For example, emerging technologies in the area of LAWS could strengthen efforts to ensure accountability over the use of force by having system logs that automatically record the operation of the weapon system. This kind of recording could facilitate investigations of both the weapon system's performance and use.

(b) Automated systems also could identify incidents meriting further review or investigation. By way of comparison, some banks, credit card companies, and other financial institutions use automated systems to identify suspicious activity and potentially fraudulent transactions. Weapons systems with autonomous functions could similarly be programmed with reporting mechanisms to highlight unusual uses meriting further review.

(c) The use of software control systems in weapon systems also creates the possibility of improving the weapon system's ability to avoid civilian casualties through updates to the software controlling the weapon system.

(d) Automated tracking systems could assist in the tracking of unexploded ordnance and fulfilling associated responsibilities under the CCW Protocol V on Explosive Remnants of War. For example, a weapon system that automatically tracked its own fires could identify and record the location where its ordnance did not explode as intended, thereby facilitating the clearance of explosive remnants of war.

[9] *Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, Oct. 23, 2018, CCW/GGE.1/2018/3.