Agenda Item 5 (c): CHARACTERISATION. The importance of critical functions

- ICRC has characterised autonomous weapon systems broadly as: “Any weapon system with autonomy in its critical functions. That is, a weapon system that can select and attack targets without human intervention.” After initial activation by a human operator, the weapon system – though its sensors, software (programming / algorithms) and connected weapon(s) – takes on the targeting functions that would normally be controlled by humans.

- Autonomy in these “critical functions” of selecting and attacking targets is central to humanitarian, legal, and ethical considerations within the scope of the Convention on Certain Conventional Weapons (CCW). It is these functions: that result in injury, damage and destruction to persons or objects in armed conflict; that are governed by international humanitarian law (IHL) rules on the conduct of hostilities; and that raise ethical questions about the role of humans in life and death decisions.

- The key distinction, in our view, from non-autonomous weapons is that the machine self-initiates an attack.

- With reference to the earlier discussion today of military applications, it may be useful to distinguish between:
  - Weapon systems where the human chooses the specific target at a particular point in time and at a specific location, for example missiles and other munitions with guidance functions; and
  - (Autonomous) weapon systems, where the exact timing, location and/or nature of the attack not known to the user since it is self-initiated by the weapon, which is – in-turn – triggered by its environment.

- As regards the latter, autonomy in critical functions is a feature that could be applied to any weapon system, especially robotic weapon systems (e.g. in the air, on land or at sea), and it is already found to a limited extent in some existing weapons, such as air defence systems, “active protection” systems and some loitering weapons.

- In the ICRC’s understanding, notions of “automated” and “autonomous” weapon systems are not easily distinguishable from a technical perspective. But more importantly, they not easily distinguishable from a humanitarian, legal and ethical perspective.

- In the ICRC’s view, the core issue is ensuring meaningful/effective/sufficient/appropriate human control over decisions to select and attack targets, independent of the technical sophistication of the weapon system.

Agenda Item 5 (d): MILITARY APPLICATIONS. Lessons for human control from autonomy in existing weapon systems

- Linking characterisation to discussions this morning on military applications, experiences with existing weapons can highlight:
  - On one hand, the potential humanitarian consequences, legal questions and ethical concerns introduced by autonomy in critical functions and;
  - On the other hand, the practical lessons for human control measures that need to be taken to mitigate these risks, and to ensure compliance with IHL, and ethical acceptability.
• The ICRC explored these issues in an international expert meeting it convened with a group of 20 States in 2016, and also proposed some elements of human control in its working paper to the CCW Meeting of High Contracting Parties in November 2018.

• In terms of lessons for human control measures, for example with existing **air defence weapons**, which have autonomous modes, States using such systems have highlighted the need for human control in three key aspects:

1) **Human supervision, and ability to intervene and deactivate**
From what the ICRC understands, and from explanations today and in previous GGE sessions with presentations by military experts, these systems are under constant supervision by a human operator with the ability – through a physical and/or communication link – to intervene and deactivate the system at any time.

   o As we understand, many such systems retain the ability, even with incoming projectiles, for a human operator to visually verify the projectile on screen and decide to cancel the attack if necessary.

2) **Predictability and reliability**
It would have to be assumed that existing systems have been tested with and meet high standards of predictability and reliability for their intended use, including the intended environment of use. In that sense, it would be beneficial for States to share the standards they use in making such assessments of predictability and reliability, including as part of national legal reviews.

3) **Operational constraints**
Operational constraints are clearly very important for human control, for increasing predictability in the consequences of use, and for compliance with IHL, in particular:

   - **Limits on the task.** These systems are used for a single, relatively simple, task to defend against incoming projectiles.
   - **Limits on the targets.** These systems are only used against *materiel* targets, such as projectiles and military aircraft, vehicles and drones. They are not used to target humans (personnel) directly.
   - **Limits on the environment.** These systems are only used in highly constrained and relatively simple environments, where there are few or no civilians or civilian objects present. As the ICRC understands, specific measures are also taken to monitor the environment and ensure civilian objects not to enter the area.
   - **Time frame of autonomous operation.** In these systems, the autonomous mode is only activated for short periods, and can be switched back to manual mode at any time.
   - **Scope of movement** over an area. These systems are fixed in place, at a perimeter, or on a ship or armoured vehicle, thereby limiting their range and effects.

• **Other existing weapons with automated or autonomous critical functions illustrate some of the difficulties** of ensuring human control and ensuring compliance with IHL. In particular, where they are unsupervised, unpredictable, or insufficiently constrained in time and space.

• For example, the use of a **loitering weapon with autonomy in critical functions** creates unpredictability for the user:

   o Due to the long time scale of autonomous operation (up to several hours) and large area of operation (up to hundreds of square kilometres), there is a high level of unpredictability
for the user as to the timing and location of the subsequent attack, which raises questions for IHL compliance.

• This unpredictability is even greater where the system is not supervised by the human operator with the ability to intervene and deactivate, should the attack need to be cancelled during that period.

• ICRC would like to hear more from military experts familiar with the operation of these types of systems.

- **Another example of a much older and simpler technology**, where unpredictability and lack of human control has raised concerns, is the mine:
  
  • The consequences of use of **anti-personnel landmines** starkly illustrate the serious dangers to civilians from weapons, which are triggered by their environment, and where the human operator has insufficient knowledge – at the point of use or activation – about the subsequent timing, (location) and nature of the attack that may result. Recognition of the resulting indiscriminate effects on civilians led to the prohibition of anti-personnel landmines by the majority of States.