IN THIS ISSUE
1 | Still in pursuit of the unizonk
5 | The right to dignity and autonomous weapon systems

11 April 2018
Vol. 6, No. 3

STILL IN PURSUIT OF THE UNIZONK
Allison Pytlak | Reaching Critical Will of Women’s International League for Peace and Freedom

In an editorial from November 2017, this publication described states’ efforts to move toward defining autonomous weapon systems as akin to pursuing a “unizonk”. This was based on a Chinese observation made in plenary, noting that some in the GGE were talking about unicorns, while others about zebras and donkeys.

It seems we are still in pursuit of the unizonk.

In the course of the thematic debate about how to characterise, or define, autonomous weapon systems, quite a lot of ideas emerged. A small group of states believe that autonomous weapon systems do not exist (the unicorn?) and there is a general sense of wanting to curtail further and future development. Donkeys, known for being stubborn, might well be a good symbol of the call from many governments to retain control. As for the zebra, well, as an animal of two colours this could represent the two functions of selecting and engaging targets which are increasingly viewed as critical.

Hybrid animal analogies aside, the discussion on Monday and Tuesday in the Group of Governmental Experts (GGE) on Lethal Autonomous Weapons (LAWs) show that there are pockets of agreement developing across several significant angles of this topic: is a working definition useful or needed? How to approach developing one? What should it include? These views are articulated in the summary below.

What is becoming confusing is that these separate aspects of discussion are sometimes bundled together in a way that leaves out some points, conflates others, with the result that states sometimes speak around and past one another. The Chair, Ambassador Gill, has helpfully produced some resources to guide us in the discussion, including a compilation of existing definitions, and often draw similarities between what he hears from governments by way of statements when providing

a summary. He has also introduced a framework for organising suggestions and approaches that is useful but it does not always align neatly with how states are organising and presenting the same information, creating a bit of a “round peg in a square hole” effect.

One new development since November is that Switzerland, Ireland, Austria, Sweden, and Canada have started to view the term “lethal” as not useful and possibly problematic. Estonia would like to consider this further, and Germany indicated not wanting to limit a definition to only autonomous weapon systems that are lethal.

Many of us in the Campaign to Stop Killer Robots have been advocating a simplified approach in order to avoid the pitfalls of dense technological discussion and using that as an excuse to stall progress. Focusing on the element of retaining meaningful human control in critical functions of identifying, selecting, and engaging targets and over individual attacks. We too do not see the benefit of only focusing on “lethal” autonomous weapons.

Thematic debate: characterisations

Thematic discussion about the “Characterisation of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the Convention” began on Monday and extended through Tuesday, including through a more interactive and informal format in the afternoon.

At the outset, Ambassador Gill, Chair of the GGE, presented what he views as the possible approaches for addressing this topic. First is what he refers to as a “separative” approach in which characteristics and concepts are identified as either not relevant, or definitely relevant, to the objectives and purposes of the CCW. Am-

continued on next page
Still in pursuit, continued

Ambassador Gill referred to these as “via negativa” and “via positiva” respectively.

Second is an “aggregative” approach, in which categories of characteristics could be added to a master-list of concepts and characteristics based on what high contracting parties bring to the discussion and then evaluated against certain criteria, such as technical, legal-humanitarian, or political-security, in order to decide on their relevance. Examples of such categories could include physical performance (mobility, speed and survivability/energy autonomy), targeting performance, and other technical characteristics or a human-machine interface related category such as system type. This approach will enable a technology or human control-centric approach.

The third approach is to avoid characterising autonomous systems altogether by using instead various levels of autonomy, other technical characteristics, or categories related to loss of human control, as benchmarks. It was suggested to refer to this as an “accountability-oriented” approach.

A fourth approach is a “purpose or effects oriented” one which states would work backward from the desired purpose to characteristics of the systems under consideration. The Chair gave the example of Mexico’s stated preference for a law-centric approach rather than a lethality-based or autonomy-based one.

Before opening for discussion, Ambassador Gill identified three caveats to bear in mind: intelligibility; not prejudging the regulatory response; and not stigmatising technology.

In its statement, China outlined five characteristics that it feels ought to be included in any definition of lethal autonomous weapons (LAWs). They include lethality; total autonomy; the impossibility of termination; having an indiscriminate nature; and the possession of self-evolutionary nature.

Belgium also listed several specific elements that might comprise a working definition. It is of the view that total autonomy in the lethal decision-making process is an important part of a definition, as are weapons where there is an unclear or uncertain division of authority between the machine and the human agent. Like China it referenced the role or existence of some form of “stop” button and expressed concern about the unpredictability of machine decision making.

Germany said that states must not view the agreement of a working definition as a goal in and of itself, and pointed out that an understanding of the desired regulatory framework is necessary to establish a working definition. Germany, noting that humans must maintain ultimate decision-making ability in matters of life and death, said it does not intend to develop or acquire weapon systems that completely exclude the human factor from decisions about the deployment of weapon systems against individuals. Its preference to determining a working definition would fall into the cumulative approach, in which autonomy and weapon systems are defined according to technical impacts or effects.

Brazil emphasised that somewhere between systems that have no automation, and those that are fully triggered, a line must be drawn via a “sound and comprehensive” working definition that will lead to an outcome, and referenced here a legally binding instrument. Brazil said that the point on the scale at which something becomes autonomous must be defined, and noted further that consideration of human-machine interaction should be central in efforts to arrive at definitions.

Bulgaria likewise sees the degree of human involvement in a system’s performance of actions such as intercepting, or attacking, as most crucial in forming definitions. It urged a focus on fully autonomous weapon systems, which, in Bulgaria’s view, do not currently exist.

The Russian Federation believes that any working definition of LAWS cannot include existing systems that have a high level of automation. It said that its advisors do not recommend defining LAWS only through the critical functions of selection of engagement, lest it send a signal that these functions can only be done by humans. It would like to avoid a breakdown of good versus bad weapons.

Italy feels that existing weapon systems governed by prescriptive rules should be considered LAWS, neither are weapon systems that have some fully autonomous functions. The debate in the current form must focus on weapon systems that are fully autonomous. It also urged taking into account the cyber-related risks to autonomous weapons systems.

Referencing its 2017 working paper, Netherlands emphasised the utility in working toward a common understanding or working definition that does not assume outcomes and is based on wider considerations than technology alone. It recommends focusing deliberations on how systems that can select and engage targets autonomously, remain under human control.

Cuba stated that it is important to define characteristics but that it is just as important to identify the aspects a linked to the implementation of relevant legal norms of international law, especially IHL. Other elements that are important to Cuba include consideration of legal and moral issues, the impact continued on next page
Still in pursuit, continued

of autonomous weapons on regional security and stability, proliferation risks, and the possibility of future arms races. It is also concerned about the use of drones and would favour an instrument that also includes these semi-autonomous weapons.

The United States delivered two interventions on Tuesday. Its first statement said again that it believes it is unnecessary for the GGE to adopt a specific working definition of LAWs without agreeing a desired political outcome, but does support identifying some general characteristics of these systems. The US feels that the four approaches outlined by the chair are helpful. In a later statement it outlined the approach of its defence department in regard to autonomous weapons development, which is not based on levels of autonomy. It warned against trying to decide what is good or bad in the abstract, or focusing on the machines above impact, both on people and law.

France says it prefers the “via negativa” approach, as it has the advantage of clarity and is immediately useful for identifying what to not include and clarify on-going misunderstandings. It said that LAWs are systems that do not exist, and would not include automated or existing remote-controlled systems.

The United Kingdom stated that the term “autonomy” is a relative one, and dependent on context. In addition, it noted that autonomy is not binary and many existing systems have functions that do, or do not, require human interaction. It also does not think that LAWs currently exist. The UK identified selection and engagement of targets as critical functions; it hopes that the concept of effective human control could be a way to focus discussion.

Ireland recommended that attempting a complex definition would be counter-productive, and “less may be better” in this context. Like others, it noted that autonomy suggests a level of independence that can exist across a spectrum and the degrees of this is at the heart of GGE discussions. Ireland takes issue with the word “lethality” being used as a qualifier and suggests removing it entirely.

Switzerland presented several ideas relating to process and substance. It referred to a definition provided in earlier working papers from 2016 and 2017 but emphasised along with others that achieving a working definition should not reflect on the final regulatory outcome. In past Switzerland has emphasised a “compliance based approach” which it still supports but also recommends a separative approach, referencing the critical functions in the targeting cycle as the most relevant to the discussions of the GGE. States should not rule out items that can change from semi- to fully autonomous, or that possess an off switch, and the degree of mobility is also unimportant.

Japan asked, at what stage should human intervention be guaranteed? It urged clarifying where legal responsibility lies. Noting that weapon systems with certain levels of autonomy have potential military advantages that should not be forgotten, Japan also explained that it has strict national standards for introducing new defense systems, which may be a good reference for GGE discussions.

The International Committee of the Red Cross (ICRC) focused its intervention around four points. The first spoke about the purpose of discussing characterisations, recommending that a broad category is a better path forward and avoids pre-judging a regulatory outcome. In the second point, the ICRC stated that autonomy in critical functions is the most important thing to consider in this discussion and agreed that lethality is not a relevant factor; to be consistent with existing law, use of force is a better concept or lens to apply. A third point is that technological sophistication is also not a defining characteristic, and lastly that autonomy does exist in both current, and potentially future, weapons.

Pakistan feels that the separative approach is most appropriate. It also warned that this discussion about definitions should be a technical issue but is becoming a political one. In its view, autonomy is the most important factor and that any system that can kill by engaging or killing without any input from humans should be considered as LAWs. It suggests considering intention as part of key characteristics.

Austria believes that technical and cumulative definitions are not helpful, and recommended determining what amount of human control is necessary. It felt that negative definitions might be helpful to narrow the subject matter, and would agree that considering meaningful human control in critical functions is a good starting point. Austria agrees with Ireland and Switzerland about use of the term “lethality”.

Estonia urged policy to drive definitions and not the other way around. It considers autonomous weapons systems as those that can engage and select targets without human involvement. Like Austria, it suggested a focus on human-machine interaction to facilitate discussion. Estonia feels that that more clarity is needed about use of the term “lethal” which it feels is a not a defining feature of any weapon system, but a de facto one. Estonia welcomes more input on this.

Kazakhstan said the need for adopting regulation on autonomous weapons is obvious. It recommended engaging information technology specialists in developing a definition, and clarified that it views

continued on next page
the issue in terms of fully autonomous weapons only. Kazakhstan said it supports creating a group of technical experts to report to the GGE about the development of LAWs and about AI.

Russia asked if the plan is changing in that states are moving to a discussion about existing autonomous and semi-autonomous weapons and ascribing to them the functions of LAWs. It feels that this approach is not fully justified and maintains that even highly autonomous systems are always under human control.

Sweden agreed with the approach of not defining end results, noting however that whatever working definition is agreed will bear some influence in how states move forward. It highlighted that there will be a spectrum of things that exist, or could exist, within a future definition of an autonomous weapon systems and some of them operate in an environment with no humanitarian impact, such as at sea. It therefore sees value in exploring further a definition that takes into account an anti-personnel versus anti-materiel specification.

Egypt stated that LAWs is an umbrella term that could include many potential and very different elements, ranging from fully autonomous weapons functioning as biological agents with free will; or weapons systems that once deployed by a human operator are not subject to human control; or weapons that once switched on could still be overridden by humans, or that remain fully controlled by humans, like drones. It said that focusing on critical functions may be distracting for the fact that we lack a definition of critical functions.

The International Committee for Robot Arms Control stressed that it will create confusion to broaden the discussion into weapons with AI or emerging AI. It feels that the ICRC definition, as concerned with autonomy in critical functions, is sufficient for the GGE’s definitional purposes. ICRAC supports those states that have stated that the focus should be on human control and human-machine interaction. Earlier on Tuesday it published a set of “Guidelines for the human control of weapons systems”.

Costa Rica, with a view to advancing the discussion on the topic, recommended avoiding an unending discussion on the characterisation of technical components. States should look at the issue of human control.

Sierra Leone, noting that since autonomy can be gradually increased or decreased, determining when a system becomes fully autonomous does not serve the interests of the GGE. It urged concentrating on level of harm, and observed that what is emerging from the discussion is that states should focus on autonomy in critical functions as basis for further work.

Panama preferred more general definitions and would like to focus on human control of critical functions, such as engagement with targets, so as to ensure inclusion of ethical and legal considerations.

Canada suggested mapping out the functions of LAWs, and the degree of autonomy allowable, as an aid to reaching a definition. The technological dynamics of “systems of systems” that must be considered.

Throughout, some delegations (France, UK, Russian Federation, Ireland, among others) referred to their existing definitions, which are contained in the compilation document.

The afternoon session moved into an interactive and more informal mode in which Ambassador Gill used the analogy of building with Lego toys for thinking about the various proposed elements and approaches of a definition and how they might be put together, or not. It was evident that some states felt more comfortable with this method than others and participation was not especially wide but succeeded in enabling a more direct and conversational dynamic particularly among those who have been most vocal on the subject. Ambassador Gill noted in his wrap-up that the approach taken matters less than the common understandings at which we arrive.

Let’s find that unizonk. ••

Special thanks to Harvard Law School’s International Human Rights Clinic for inputs to this summary. This summary is not meant to be exhaustive of all statements delivered.
All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

— Article 1 of the Universal Declaration of Human Rights

Since the Convention on Conventional Weapons (CCW) began exploring issues surrounding autonomous weapon systems (AWS) in 2014, much attention has been paid to potential violations of international humanitarian law and strategic challenges for their use by militaries. Less attention has been paid to the threat AWS pose to core principles of international human rights norms. When examined, however, it is apparent that the introduction of AWS into military arsenals could violate several well-established human rights norms. When examined, however, it is apparent that the introduction of AWS into military arsenals could violate several well-established human rights norms such as the right to dignity as well as others including the right to life and the right to remedy. For example, the potential erosion of human participation in warfare by the use of AWS will rob its victims of their right to dignity in two key ways: by eliminating responsibility and by allowing someone’s life to be taken by an algorithm.

In any conflict, human beings have responsibilities associated with warfare. Historic developments in military equipment have allowed forces to further distance themselves from the battlefield, as well as from the consequences of their decisions. To date, none of these developments have completely removed human control and responsibility from warfare, as AWS threaten to do. Contemporary warfare still requires combatants to assume some degree of responsibility, while AWS could allow operators to avoid legal, emotional, and ethical responsibility in warfare.

A human being is capable of empathising with a targeted individual no matter the distance and is aware that they will have to live with the legal, emotional, and ethical consequences of their decision. As a result, a human being will seek to justify her or her own decision before firing. AWS, which could not have empathy, would not be capable of considering such consequences. Deploying AWS in combat displays the belief that any human targeted in this way does not warrant the consideration of a live operator, thereby robbing that human life of its right to dignity.

AWS can further erode targets of their right to dignity by replacing them with an algorithm. Programming a computer to decide who lives and dies by reducing a person to a series of coded physical characteristics, potentially perceiving an increased threat level from specific ethnicities, genders, ages, and religions, creates a scenario in which violations of international law are a near certainty. Furthermore, the use of AWS with machine-learning capabilities presents an even greater threat to the right to dignity; once capable of deducing patterns from operations in the field, such machines are likely to develop a stereotypical enemy combatant that would result in discrimination. Such trends are already apparent in contemporary artificial intelligence. For example, a police force in the United Kingdom using artificial intelligence to determine the likelihood of repeat offenders found that the software was discriminating against those living in low income areas.

This week at the Group of Governmental Experts (GGE) meeting, states should ensure that human rights norms, including the right to dignity, are not overlooked. International human rights law applies at all times, including during armed conflict, and thus is not an optional addition to international humanitarian law when discussing weapons of war. Furthermore, international human rights law applies to law enforcement situations in which AWS might also be used. Given that AWS would operate without human control, it is highly unlikely that AWS could comply with states’ responsibility towards the right to dignity. States should therefore call for an obligation to maintain meaningful human control over the use of force and a pre-emptive ban on the development, production, and use of weapons that lack such control.

Notes
