Statement by iPRAW during the CCW GGE on LAWS: Human Control

Delivered by Marcel Dickow (Twitter: @MarcelDickow) on March 29th, 2019

Thank you Mr. Chair!

Building on the work of the International Panel on the Regulation of Autonomous Weapons of the past two years, I would like to present a few observations and questions regarding the human element in the use of force in a more conceptual way. I hope States Parties will find this useful for their deliberations.

We see a majority of States Parties agreeing on the relevance of human control as a concept and would like to emphasize the importance of this approach in contrast to definitions of **either autonomy or automation**. I will give an example to illustrate this: Imagine two weapon systems, both performing the selection and engagement of targets without human intervention. The first system intercepts incoming munitions. The second system additionally counterattacks against the source of incoming munition. In the first case, firing at incoming munitions, meaningful human control could be implemented through the design of the systems and continued supervision after its activation. In contrast, the second situation, the counterattack, would require an understanding of the specific situation at the source of the enemy fire and a legal judgement prior to the use of force. So, we see: Despite the functionality being *the same* in both cases, context matters, and thus different legal judgements and different types of human control are required. So it is only this focus on the type of human control that allows for conceptually grasping the problem. Semantics, such as attempts to delineate automation from autonomy, are not only *not* helpful but are in fact running the risk of becoming a distraction within the debate. We therefore emphasize the importance to set the focus of the GGE's deliberations on the description of minimum requirements for human control.

iPRAW's **concept of human control** includes the situational understanding and options for intervention in design and use of the system. Those and similar terms have been used over course of the past few days. While iPRAW usually refers to a human operator or commander, the concept can be expanded to distributed authority just as well.

Situational understanding means that the human operator is aware of the environment and the mode of the system during the operation. The awareness regarding the environment is necessary because battlespace situations change, for instance if civilians enter or if a combatant surrenders or is wounded and thus *hors de combat*. The supervision of the

system itself is important to discover malfunctions or hacking before a catastrophic effect occurs. This influences both the system's internal design and interface.

Situational understanding as a dynamic concept depends on context and application and may vary in quality and quantity even within a given system. The operator does not have to be able to understand the system on a software level. Nevertheless, the system's design and interface must allow the operator to understand why the system has produced a specific outcome. Especially the design of failure modes in all stages of the targeting cycle must allow for enough time and information for situational understanding. That would include a clear indication of responsibilities (What is demanded from the operator?) and an immediate halt on the use of force.

One aspect that has been mentioned in relation to autonomy is enhanced **precision**, including the notion that more human control could lower the targeting precision in certain situations. We would like to encourage States to present examples for enhanced precision through autonomy not only in the *effect* of the weapon but in the target *selection*. In our view the term precision relates to the weapon's physical effect, in particular a weapon system's spatial and temporal ability to hit the pre-defined target. We also sensed an interpretation of the term precision to previous steps within the targeting cycle including the selection of targets. IPRAW sees a need for further clarification between the link of an absence of human decision making in the targeting steps and precision in the selection of targets.

Autonomous functions do not preclude the application of human control during operation. On the contrary, novel techniques to ensure military efficiency often allow for better human control due to increased and more frequent updates for situational understanding and a possibility for timelier intervention. In our understanding this *cooperation* of human and machine *may* lead to higher precision of weapon systems.

With regard to Paragraph 7 of Agenda Item 5a and Paragraph 5 of Agenda Item 5c, I would like to share some observations from iPRAW's discussions on the concept of constrains of weapon systems with autonomous functions in time and space, or as we called it "boxed autonomy": the predictability of a system is a technical feature. These constrains may enable the commander or operator of a weapon system to mitigate problematic effects of autonomous functioning within the box – but they do not enhance the predictability of the system per se.

The final aspect I would like to comment on are the regulative **options** to strengthen or implement the broadly endorsed principle of human control. In our recent report, we come to the conclusion that inaction is not a viable option: autonomous weapon systems raise fundamental questions and the related ethical, legal, and political implications are too wide-ranging and important to remain unaddressed. National weapon reviews alone are not sufficient to address those issues. iPRAW considers it important for States Parties to take regulatory action to shape whether and how LAWS are developed. Human control has to be the foundation of any policy that is formulated.

Thank you!