

## Focus on Human Control

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

### Empfohlene Zitierung / Suggested Citation:

International Panel on the Regulation of Autonomous Weapons (iPRAW). (2019). *Focus on Human Control*. ("Focus on" Report, 5). . <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-77410-6>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



# Focus on Human Control

“Focus on” Report No. 5

August 2019

International Panel on the Regulation of Autonomous Weapons (IPRAW)

coordinated by:

Stiftung Wissenschaft und Politik (SWP) – German Institute for International and Security Affairs  
Ludwigkirchplatz 3-4  
10719 Berlin, Germany

August 2019

[www.ipraw.org](http://www.ipraw.org)  
[mail@ipraw.org](mailto:mail@ipraw.org)

This project is financially supported by the German Federal Foreign Office.

## ABOUT IPRAW

**Setting and Objectives:** The International Panel on the Regulation of Autonomous Weapons (iPRAW) was founded in March 2017. iPRAW is an independent group of experts from different nation states and scientific backgrounds. The panel will complete its work by the end of 2019. The mission of iPRAW is to provide an independent source of information and consultation to the Group of Governmental Experts (GGE) within the framework of the United Nations CCW (Convention on Certain Conventional Weapons) during the ongoing process toward a possible future regulation of LAWS (Lethal Autonomous Weapon Systems). This work includes, but is not limited to, the provision of expertise on the military, technical, legal, and ethical basis for **practical and achievable policy initiatives regarding LAWS. The mandate of the CCW's open-ended GGE on LAWS will guide the work of iPRAW.** iPRAW seeks to prepare, support, and foster a frank and productive exchange among participants, culminating in perspectives on working definitions and recommendations on a potential regulation of LAWS for the CCW GGE. iPRAW is independent from the GGE and does not function in any official capacity regarding the CCW.

**Funding, Organization, and Participants:** iPRAW is financially supported by the German Federal Foreign Office. The views and findings of iPRAW do not reflect the official positions of the German government or any other government. Stiftung Wissenschaft und Politik – The German Institute for International and Security Affairs (SWP) organizes the panel. The participants have been selected on the basis of their expertise and the perspectives they bring from a wide range of professional and regional contexts. iPRAW represents the diversity of views on the topic of autonomy in weapon systems. Its members have backgrounds in natural science, engineering, law, ethics, political science, and military operational analysis.

**Scope:** The panel acknowledges that LAWS may pose a number of considerable legal, ethical and operational challenges and that they might change the security environment in a fundamental way. The full potential of these weapon systems is yet unknown and a mutually agreed definition on LAWS does not exist. In order to support the CCW GGE process, iPRAW will work on **possible approaches to a regulation of LAWS. The panel's working session will cover the following topics**

- Requirements for human control over the use of force
- Verification of a potential regulation of LAWS
- Proliferation and export control

iPRAW will publish a respective report aimed at informing the CCW process.

**Procedure:** The participants commit themselves to actively participate in and contribute to the meeting and the scientific dialogue related to iPRAW's activities. Papers with agreed upon recommendations on relevant issues will be drafted and published via the project's website ([www.ipraw.org](http://www.ipraw.org)).

**Communication and Publication:** The participants discuss under the Chatham House Rule: participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed. As a matter of confidentiality, photographs, video or audio recordings as well as all kinds of activities on social media are not allowed during iPRAW meetings. The results of the panel discussions will be published. iPRAW members will strive to reach consensus on their recommendations and to reflect that in the **panel's publications. Media inquiries with regard to official iPRAW positions should be directed to the steering group.** Apart from that, the panel members are free to talk about their personal views on participation and the topics of the panel.

Learn more about iPRAW and its research topics on [www.ipraw.org](http://www.ipraw.org). Please direct your questions and remarks about the project to [mail@ipraw.org](mailto:mail@ipraw.org).

# CONTENT

- Executive Summary.....5
- 1 Introduction .....7
- 2 Considerations on Human Control .....8
  - 2.1 Concepts of Human Control in the Use of Force.....8
  - 2.2 The Relation of Human Control and IHL .....10
  - 2.3 **iPRAW’s Approach to Human Control**.....12
- 3 The (Challenge of Developing a) Standard for Human Control .....14
  - 3.1 A More Precisely Defined Notion of Attack?.....14
  - 3.2 The Operational Context .....15
- 4 Conclusion .....17
- 5 Annex .....19
  - 5.1 Literature.....19
  - 5.2 Members of iPRAW .....21

## EXECUTIVE SUMMARY

The International Panel on the Regulation of Autonomous Weapons (iPRAW) is an independent, interdisciplinary group of scientists working on the issue of lethal autonomous weapon systems (LAWS). It aims to support the current debate within the UN Convention on Certain Conventional Weapons (CCW) with scientifically grounded information and recommendations for the potential regulation of LAWS. Defining LAWS is a critical element of the CCW debate and as such a major **component of iPRAW's mission**. iPRAW publishes interim reports that each focus on different aspects or perspectives on LAWS. This report focuses on human control in the use of force, picking up questions that iPRAW has not addressed in its first installment in 2017-18. Following the observations stated in the report, iPRAW makes the following conclusions for aspects of a potential regulation of LAWS:

In this report, we picked up considerations on human control from previous iPRAW reports, linked them to IHL requirements and identified remaining conceptual questions. In particular, we discussed requirements for human control from IHL, the **link between human decision making and the IHL concept of 'attack', and the influence of the operational context on the implementation of human control**.

**Human Control as a Consequence of IHL:** Autonomous functions in weapon systems call for human control before and during the attack. Precaution during attack remains feasible when the operator/commander has sufficient situational **understanding and options for intervention along the lines of iPRAW's concept of human control**. Accordingly, the operator/commander must be enabled to review **legal assessments and translate human decision making into the system's action** during attack prior to the actual engagement.

**More Precise Notion of Attack:** Defining what constitutes the start of an attack can be useful in unpacking the concept of human control. The most relevant point in the mission thread is not defined by the launch or activation, but by the final necessary decision on target selection and engagement by a human. Weapon systems with autonomous functions potentially move the final human decision to a very early stage of the operation. With regard to the legal judgments to abide by IHL principles this effect could be challenging for two reasons: First, it can increase the level of abstraction in the target selection process (i.e. class of targets instead of specific

target). Second, the environment might change during this extended timespan between targeting decision and engagement, e.g. outdated the initial proportionality assessments.

The underlying notion of attack will therefore influence the understanding of the principle of human control in a regulation of autonomous weapon systems. This is because IHL principles like distinction and proportionality are legally required during the planning phase, but, to a certain extent, become a question of feasibility in attack. This would alter the need or necessary level of human control in attack.

**Context-Dependency of Human Control:** While it is possible to develop abstract minimum requirements for human control in the use of force, the appropriate level or implementation of human control depends on the details of the operational context. A **'one-size-of-control-fits-all' solution** that addresses all concerns raised by the use of autonomous weapon systems will most likely not be achievable because it cannot account for the multitude of combinations of environmental factors, operational requirements, and weapons capabilities. Instead a (binding or non-binding) regulation would be more useful if it included general approximations to be specified in each case along the lines of existing IHL considerations. iPRAW encourages CCW States Parties to develop and share specific examples for how control by design and control in use can be implemented in weapon systems used in different operational contexts.

## 1 INTRODUCTION

This report builds on iPRAW's previous work and elaborates on further details regarding the human role in the use of force. It focuses on three strands of the debate: the basis of human control in international humanitarian law (IHL), the notion of the legal term 'attack', and the adequate level of human control in relation to the operational context – and examines how they are interrelated. While these are mostly legal issues, they are also closely linked to operational considerations.

In Chapter 2, we first frame some of the open questions regarding human control within the context of existing concepts of human control in the use of force, analyze how human control relates to IHL, and link iPRAW's own approach to the debate. While iPRAW's approach towards human control aspires to complement existing concepts, it leaves some open questions itself though. For example, it does not address the notion of 'attack' and it does not conceptualize the implications of the operational context. The particular time span (begin and end) of an attack is relevant to defining the appropriate implementation of human control because elements of the IHL principle of precaution must be applied *during* attack. Those aspects of human control will guide our deliberation in this report and will be further discussed in Chapter 3.

**Why focusing on human control?** Although much attention has centered on how to define autonomy, this focus did not advance the debate on why a definition matters or how it impacts the option of a regulation. In the course of iPRAW's analysis, the limitations of this focus on definitions has emerged. As a result, iPRAW decided to focus on human control as the key concept to understand for the CCW debate on LAWS, and as the foundation for the range of regulatory options available to the CCW. The question of human control brings different aspects of the human-machine relation to the fore in a way that the debate on definitions could not. It is essential to understand that control is a context-dependent term and this is precisely why it is useful as an anchor concept: it forces one to consider the variables of the environment and the human-machine relationship and how they may be impacted with differing applications of machine autonomy.



## 2 CONSIDERATIONS ON HUMAN CONTROL

In our previous report we pointed out the practical limitations which we discovered in defining autonomy as a means to regulate LAWS.<sup>1</sup> As a result, our investigation focuses on human control as indispensable concept to accelerate the CCW debate on LAWS. In this chapter, we first highlight a few existing approaches to conceptualize human control in the use of force, discuss the impact of IHL on the interpretation of human control, and examine which aspects of human control iPROAW has covered so far.

### 2.1 CONCEPTS OF HUMAN CONTROL IN THE USE OF FORCE

Various concepts have been influencing iPROAW's thinking about human control. In the following we will highlight different important aspects with their relation to IHL, in particular the use of force.

**Layers of the use of force:** In 2016 Roff and Moyes presented the first conceptual approach towards (meaningful) human control. They assert that human control is enhanced if a) the *weapon system* is predictable, reliable, and transparent, and b) the *user* has accurate information as well as the ability for timely human action and accountability. In this view, human control comprises a technical design element and **specific requirements for the human**. Roff and Moyes divide the 'use of force' into three layers: (1) design, development, acquisition and training, (2) attack (as used in IHL and following a phase of operational or strategic planning), and (3) command structures and accountability. According to Roff and Moyes, human control has to be exerted on the lowest possible level during an attack. This notion of human control is understood as a requirement from IHL, because "humans are the agents that a party

<sup>1</sup> See International Panel on the Regulation of Autonomous Weapons (December 2018), *Concluding Report*, <[https://www.ipraw.org/wp-content/uploads/2018/12/2018-12-14\\_iPROAW\\_Concluding-Report.pdf](https://www.ipraw.org/wp-content/uploads/2018/12/2018-12-14_iPROAW_Concluding-Report.pdf)> (August 08, 2019).

to a conflict relies upon to engage in hostilities, and are the addressees of the law as written”<sup>2</sup>.

While offering extremely helpful conceptual input with regard to different phases of human control, the paper does not address the influence of IHL-relevant judgments in the planning phase on the execution of the attack, which becomes especially relevant in deliberate targeting.<sup>3</sup> It therefore remains unclear at which point in time human control – in the form of target selection and final decision making – needs to ultimately take place. The question about when an attack begins (and ends) is of specific legal relevance due to different obligations before and during attack – leading to differing legal interpretations of human control.

**Modes of Target Selection:** Sauer et al. built on the analyses by Roff and Moyes and developed a taxonomy for describing different levels of control of LAWS, similar to levels of autonomy that exist in the automotive industry.<sup>4</sup> To define acceptable modes of target selection, Sauer et al. suggest the following five levels of human supervisory control: “(1) human deliberates about a target before initiating any and every attack; (2) software provides a list of targets and human chooses which to attack; (3) software selects target and human must approve before attack; (4) software selects target and human has restricted time to veto; (5) software selects target and initiates attack without human involvement”<sup>5</sup>. According to this approach level 4 and 5 are highly problematic because the human lacks situational understanding and/or adequate options for intervention. Those levels would only be acceptable in the very restricted context of defensive, purely anti-materiel purposes.

<sup>2</sup> Heather Roff & Richard Moyes (April 2016), *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons*, <<http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>> (August 08, 2019), p. 5.

<sup>3</sup> So far, iPROAW focused on the use of force in the (United States’ military) dynamic targeting cycle as a framework for analysis. In that context, dynamic targeting is defined as a time compressed process of finding, fixing, tracking, targeting, and engaging the target and assessing the outcome (F2T2EA). Deliberate targeting is based on similar steps, but allows for more preparation and planning because engagement can be scheduled.

For similar considerations on NATO procedures see: NATO (April 2016), *NATO Standard AJP-3.9. Allied Joint Doctrine for Joint Targeting*, <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/628215/20160505-nato\\_targeting\\_ajp\\_3\\_9.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/628215/20160505-nato_targeting_ajp_3_9.pdf)> (August 08, 2019), p. 1-2 - 1-3; for the application of the deliberate NATO procedures to autonomous functions in weapon systems, see: Merel Ekelhof (2018), *Lifting the Fog of Targeting: “Autonomous Weapons” and Human Control through the Lens of Military Targeting*, in: *Naval War College Review*: (71) No. 3, Article 6.

<sup>4</sup> See e.g. Society of Automotive Engineers (September 2016), *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016\_201609)*, <[https://www.sae.org/standards/content/j3016\\_201609/](https://www.sae.org/standards/content/j3016_201609/)> (August 08, 2019).

<sup>5</sup> Frank Sauer et al. (2018), *Autonomy in Weapon Systems. The Military Application of Artificial Intelligence as a Litmus Test for Germany’s New Foreign and Security Policy*, Heinrich Böll Foundation, <[https://www.boell.de/sites/default/files/boell\\_autonomy-in-weapon-systems\\_v04\\_kommentierbar\\_1.pdf?dimension1=division\\_oen](https://www.boell.de/sites/default/files/boell_autonomy-in-weapon-systems_v04_kommentierbar_1.pdf?dimension1=division_oen)> (August 08, 2019), p. 42.

This classification is very helpful in breaking down different modes of target selection and evaluating the appropriateness of different levels of human control to the context. We discuss the issue of context-dependency further in Chapter 3. Since this approach focuses on the human-machine relation within the targeting process, it does not define the notion of attack, i.e. what is necessary to start an attack and what role the human plays in that.

## 2.2 THE RELATION OF HUMAN CONTROL AND IHL

It is disputed if the need for human control during an attack can be derived from IHL. One key question in this discussion is if it is necessary for the user to select a specific target to make legal judgments according to IHL principles. In the use of existing long-distance stand-off weapons (e.g. cruise missiles, torpedoes or artillery) a human usually defines *one* specific target through a combination of signature, location and/or trajectory.<sup>6</sup> In contrast to that, weapon systems with autonomous functions would allow for more abstract targeting criteria, e.g. *classes* of targets. This distinction is relevant for legal judgments of IHL because the specific target selection during attack is left to the machine. Therefore the machine would have to ‘calculate’ the concrete (e.g. proportionality) assessment – which is highly context-dependent and subjective.

In addition to this, the assessment of IHL principles during attack in order to avoid civilian casualties is subject to feasibility as laid down in the principle of feasible precautions, which is enshrined in customary IHL:

**“In the conduct of military operations, constant care must be taken to spare the civilian population, civilians and civilian objects. All *feasible* precautions must be taken to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians and damage to civilian objects.”<sup>7</sup>**

While the commander must consider IHL principles, such as distinction and proportionality during the planning phase (e.g. choice of ammunition, time and location of the engagement), the application of those requirements during the attack depends on feasibility.

The combination of abstract targeting criteria in weapon systems with autonomous functions and the limited IHL standards during attack could lead to two different legal interpretations regarding human control in the use of force:

<sup>6</sup> Even in the use of those weapons the timespan between decision and engagement can become extended. This can be problematic with regard to dynamic changes on the battlefield and environment but is usually accepted. Novel technologies in weapon systems provide communication links to enable the operator to terminate the mission or re-task the weapon system even after launch.

In addition the selection and engagement of a specific target by solely a machine has been a niche capability so far and is only now emerging as a potentially wide-spread autonomous function in weapon systems.

<sup>7</sup> International Committee of the Red Cross, *Rule 15. Precautions in Attack*, IHL Database – Customary IHL, <[https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_rul\\_rule15](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule15)> (August 08, 2019). Similar in: Article 57 Additional Protocol I to the Geneva Conventions.

**Approach 1: “the reliable machine” where human control during attack is not required per se:** According to this perspective, IHL does not necessarily call for human control during an attack as long as the commander’s intention is translated into the desired effects on the battlefield. The means and measures are the commander’s choice and depend on her evaluation. They must include considerations on distinction, military necessity, proportionality, and feasible precautions, which require that the choice of weapons/munition is adequate to the expected circumstances and that the commander has the responsibility to account for dynamic changes in the environment.

In this interpretation, the commander would need a sufficient situational understanding of the environment before the attack begins. Furthermore, the machine’s action in fulfilling the given task must be predictable and reliable, assuming that the dynamics of the environment are predictable, too. Beyond that, human control over the machine during the attack depends on the operational goal and feasibility.

**Approach 2: “a human touch” where human control during attack is required, and adjustments to IHL are necessary:** According to a second perspective, the principle of precaution inherently requires some sort of human control during the attack because the abstract target selection criteria do not allow for sufficient legal assessments prior to the attack. This is based on the assumption that legal decisions including proportionality assessments and the definition of the military objective require an understanding of the context and an interpretation through human cognition<sup>8</sup>. Thus even if the commander controls the temporal and spatial scope of the attack (‘kill box’), she would be obligated to supervise the attack and would need an option to cancel the attack at any point if circumstances changed. This stipulates a norm of human control during attack.

Since IHL as of yet does not say anything specific about the acceptable levels of unpredictability or testing, about details of human control or about the possibility of a kill box, it seems prudent to clarify and future-proof the existing law in that regard.

**Conclusion:** Among other things, autonomous functions in weapon systems allow the operator/commander to delegate the actual selection of a target to the machine by using classes of targets rather than one specific target. While Approach 2 requires the operator/commander to have situational understanding and options of intervention during attack, Approach 1 calls for a lower standard of human control

<sup>8</sup> See International Committee of the Red Cross (April 2018), *Statement at the CCW GGE on LAWS: Further consideration of the human element in the use of lethal force; aspects of human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems*, <[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/5216D20D2E98E7AAC12582720057E6FC/\\$file/2018\\_LAWS6b\\_ICRC1.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/5216D20D2E98E7AAC12582720057E6FC/$file/2018_LAWS6b_ICRC1.pdf)> (August 08, 2019): “The ICRC is clear that the law is addressed to States and humans, and the legal obligations under international humanitarian law ultimately rest with combatants who plan, decide upon, and carry out attacks. Combatants will require a minimum level of human control over weapon systems with autonomy in their critical functions so that they can effectively make legal judgements – of distinction, proportionality and precautions – in specific attacks. Human control can take different forms during the development, activation, and operation of an autonomous weapon system. However, these legal judgements are context specific. Therefore, concerns will arise where the design and/or use of the weapon interferes with combatants’ ability to make the necessary legal judgements in carrying out attacks.”

during attack. Subsequently the beginning of the attack is of crucial importance for the legal assessment in accordance with IHL. When the final human action required to trigger the use of force can become so detached from the effect with regard to time, space, and the actual target eventually selected, it might be necessary to spell out more explicitly and clearly the notion of attack, in particular its beginning (see Chapter 3.1).

### 2.3 IPRAW’S APPROACH TO HUMAN CONTROL

The second approach presented above is in line with iPRAW’s perspective on human control, which calls for commanders/operators to have a sufficient and up-to-date situational understanding of an attack and options for intervention by design and in use. That means that the design of weapon systems with autonomous functions must enable the operator/commander to *understand* the operational context to allow for informed decisions over each step of the use of force.<sup>9</sup> The necessary monitoring of the environment and the system includes system diagnostics, internal and external sensors for system and environmental monitoring as well as methods for communicating that information. In addition, the ability for humans to at any time actively *intervene* prior to the ultimate use of force should be a default feature.

The need for situational understanding and intervention is not limited to one single weapon system, but should also refer to systems of multiple robots executing a shared mission, which is how these capabilities will likely be developed and fielded.

	Situational Understanding	Intervention
<b>Control by Design</b> (Technical Control)	Design of systems that allows human commanders the ability to monitor information about environment and system	Design of systems with modes of operation that allow human intervention and require their input in specific steps of the targeting cycle based on their situational understanding
<b>Control in Use</b> (Operational Control)	Appropriate monitoring of the system and the operational environment	Authority and accountability of human operators, teammates and commanders; abide by IHL

Table 1: Requirements for Human Control in the Use of Force

iPRAW’s approach to human control is focused on the **necessity of humans to be making targeting decisions**, including both technical (i.e. the design of a weapon system: control by design) and operational requirements (i.e. the procedures to maintain control over the weapon systems: control in use). Both incorporate measures earlier in the life cycle of a weapon system to ensure that the need for human control in operation is considered during research and development, programming and testing of systems, and deployment of various components of a LAWS. While responsible innovation and research is a key element to shaping the thinking of developers with regard to compliance with relevant legal frameworks, it is the responsibility of states to ensure that their military requirements necessitate

<sup>9</sup> The following paragraphs are taken from iPRAW’s Concluding Report, see Fn. 1.

human control for the development of new weapon systems. This particularly includes the commercial sector.

The implementation of military objectives is based on design features and capabilities of weapon systems combined in a procedure called mission planning. Even though this is a crucial step in a military operation, iPRAW's concept of human control would consider it as insufficient if the human involvement was limited solely to the planning phase of the mission – especially with **Approach 2 (“The human touch”)** in mind. This also aligns with the caution regarding ‘boxed autonomy’ that iPRAW expressed in previous reports. Similarly, Sauer et al. ruled out their Levels 4 and 5 in most cases because those preserve human decision making on target selection only in the mission planning phase.

Autonomous functions enable the weapon system to potentially proceed through different steps of the targeting cycle without human interference. When the weapon system is activated, it can act and eventually engage targets at a later point and different place without a human operator/commander in or on the decision loop. It therefore spans a time period and geographical space of autonomous functioning, or as iPRAW calls it, **boxed autonomy**. Without human supervision (situational understanding) and options for intervention, this box is an extension of initial human judgment and decision making over time and space. Limiting the box to predefined parameters, a fixed time period and geographical borders, is a necessary element of safeguarding human control over the weapon system as it constrains the unpredictability of the environmental factors. It neither directly increases the predictability of the system itself nor the predictability of the operational factors. The latter ones are to some extent dependent on adversarial behavior and can change even in narrow box parameters. iPRAW is thus of the opinion that boxed autonomy can only be implemented in an IHL compliant fashion when control by design and control in use strictly limit its application to, for example, defense against incoming munitions.

iPRAW has not explicitly applied its concept of human control against the background of **distributed authority** over a (complex) weapon system. We never excluded, however, that situational understanding and options for intervention might be distributed within the chain of command and vary from authority to authority. We also explicitly referred to the use of multiple heterogeneous systems and sensor data fusing.<sup>10</sup>

<sup>10</sup> See International Panel on the Regulation of Autonomous Weapons (March 2018), *Focus on the Human-Machine Relation in LAWS*, <[https://www.ipraw.org/wp-content/uploads/2018/03/2018-03-29\\_iPRAW\\_Focus-On-Report-3.pdf](https://www.ipraw.org/wp-content/uploads/2018/03/2018-03-29_iPRAW_Focus-On-Report-3.pdf)> (August 08, 2019), p. 15.

### 3 THE (CHALLENGE OF DEVELOPING A) STANDARD FOR HUMAN CONTROL

This chapter sheds some light on two remaining issues: the notion of attack and the role of the operational context. Human control, as argued in previous iPROW reports, is a context-dependent concept – the range of variables construing the context of a situation calls for flexible and dynamic ways of human control.

#### 3.1 A MORE PRECISELY DEFINED NOTION OF ATTACK?

What marks the beginning of an attack? As mentioned above, defining what constitutes the start of an attack can be useful in unpacking the concept of human control. However, the timeframe of an ‘**attack**’<sup>11</sup> is not defined by IHL even though the extent of an attack in time and space has an effect on the legal obligations. In general, an attack can be understood as the application of force, such as killing or injuring humans and destroying objects. However, its starting point is ambiguous when we consider weapons that travel long distances (e.g. missiles) or can remain inactive for a long time (e.g. loitering munitions).

As shown above, the intended and expected effect of the lethal force initiated is determined and programmed way before the destructive impact takes place. Therefore the beginning of the attack with regard to weapons like loitering munitions can be understood as the point in time when the munition becomes a danger to people<sup>12</sup> – which is usually its launch. It could also be its remote activation depending on the necessity for the operator to intervene after its launch. Those steps can be seen as elements of a mission thread. A mission thread is a sequence of tasks performed to accomplish a mission in a given scenario. Here it serves as a helpful concept to identify steps with possible human involvement in a military operation.

<sup>11</sup> Art. 49(1) Additional Protocol I to the Geneva Conventions **defines an attacks as** “acts of violence against the adversary, whether in offence or in defence”.

<sup>12</sup> At least the commentary to Art. 49(1) Additional Protocol I to the Geneva Conventions indicates **this understanding with regard to mines**: “the question arose whether the placing of mines constituted an attack. The general feeling was that there is an attack whenever a person is directly endangered by a mine laid.”

The figure below illustrates (very simplified) how the begin of the attack would change depending on the final human decision that sets the use of force into motion.

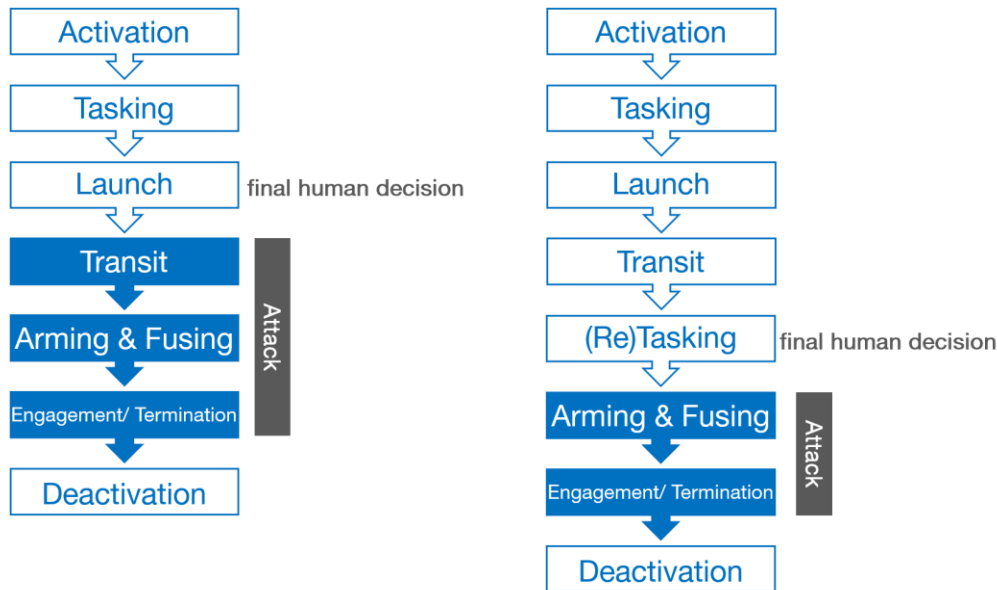


Figure 1: Two hypothetical mission threads of operations with autonomous weapons

This points to an inherent consequence of modern weapon technologies with autonomous functions: the final necessary decision on target selection and engagement by a human becomes more relevant to define the beginning of an attack than terms like ‘launch’ or ‘activation’.

It is iPRAW’s understanding that human control during each attack requires frequent situational understanding and the technical and procedural option for human intervention to allow for a continued application of the principles of distinction, proportionality and precaution.<sup>13</sup>

### 3.2 THE OPERATIONAL CONTEXT

The operational context is crucial for defining the necessary type and level of human control;<sup>14</sup> all the more since multiple factors contribute to the determination of what level of human control is adequate in a given situation. A ‘one-size-of-control-fits-all’ solution that addresses all concerns raised by the use of autonomous weapon systems will thus most likely not be achievable.

Looking at this multitude of relevant factors from the perspective of IHL, the crucial lens would be the **risk for violations of IHL** (due to a lack of situational understanding or timely intervention), i.e. if combatants and civilians cannot be distinguished properly or if the proportionally assessment cannot be made adequately. **One factor contributing to this is the predictability of the**

<sup>13</sup> If States wanted to keep the beginning of the attack as close as possible to the effect, they could require – instead of just the option for intervention – an active decision and action by a human operator to pick and engage a target (assuming that the operator is capable of an informed decision).

<sup>14</sup> Not: the need for human control.



environment, e.g. the likeliness for changes with potential for civilians or hors de combat ('dynamic, cluttered environment'). Sauer et al. add some examples for contextual factors relevant to this assessment: the geographical range of the weapon, the purpose, and the type of target. In certain circumstances, e.g. a very limited range due to a stationary system, the deliberate or physical technical limitation of the used munitions, the domain of application, and a purely anti-materiel weapon, the necessary degree of human control can be substantially lower than in others.<sup>15</sup>

During its meeting in May 2019, iPROAW discussed a possible classification of factors that define the operational context in order to derive consequences for the implementation of human control.<sup>16</sup> We dismissed this kind of typology because it cannot account for the multitude of combinations of environmental factors, operational requirements, and weapon capabilities. We rather invite states to present their solutions for the implementation of human control in specific applications. That would allow participants to the CCW GGE to learn more about the appropriate level of human control.

<sup>15</sup> See Sauer et al., Fn. 5.

<sup>16</sup> For an exploratory approach to define relevant criteria see: Marcel Dickow et al. (2015), *First Steps towards a Multidimensional Autonomy Risk Assessment (MARA) in Weapons Systems*, <[https://www.swp-berlin.org/fileadmin/contents/products/arbeitspapiere/FG03\\_WP05\\_2015\\_MARA.pdf](https://www.swp-berlin.org/fileadmin/contents/products/arbeitspapiere/FG03_WP05_2015_MARA.pdf)> (August 08, 2019).

## 4 CONCLUSION

In this report, we picked up considerations on human control from previous iPRAW reports, linked them to IHL requirements and identified remaining conceptual questions. In particular, we discussed requirements for human control from IHL, the link between human decision making and the IHL concept of ‘attack’, and the influence of the operational context on the implementation of human control.

**Human Control as a Consequence of IHL:** Autonomous functions in weapon systems call for human control before and during the attack. Precaution during attack remains feasible when the operator/commander has sufficient situational understanding and options for intervention along the lines of iPRAW’s concept of human control. Accordingly, the operator/commander must be enabled to review legal assessments and translate human decision making into the system’s action during attack prior to the actual engagement. The above mentioned Approach 2 (“the human touch”) could serve as a starting point to interpret IHL in accordance with iPRAW’s concept of human control.

**More Precise Notion of Attack:** Defining what constitutes the start of an attack can be useful in unpacking the concept of human control. The most relevant point in the mission thread is not defined by the launch or activation, but by the final necessary decision on target selection and engagement by a human. Weapon systems with autonomous functions potentially move the final human decision to a very early stage of the operation. With regard to the legal judgments to abide by IHL principles this effect could be challenging for two reasons: First, it can increase the level of abstraction in the target selection process (i.e. class of targets instead of specific target). Second, the environment might change during this extended timespan between targeting decision and engagement, e.g. outdated the initial proportionality assessments.

The underlying notion of attack will therefore influence the understanding of the principle of human control in a regulation of autonomous weapon systems. This is because IHL principles like distinction and proportionality are legally required during the planning phase, but, to a certain extent, become a question of feasibility in attack. This would alter the need or necessary level of human control in attack.

**Context-Dependency of Human Control:** While it is possible to develop abstract minimum requirements for human control in the use of force, the appropriate level or implementation of human control depends on the details of the operational context. A **'one-size-of-control-fits-all' solution** that addresses all concerns raised by the use of autonomous weapon systems will most likely not be achievable because it cannot account for the multitude of combinations of environmental factors, operational requirements, and weapons capabilities. Instead a (binding or non-binding) regulation would be more useful if it included general approximations to be specified in each case along the lines of existing IHL considerations. iPRAW encourages CCW States Parties to develop and share specific examples for how control by design and control in use can be implemented in weapon systems used in different operational contexts.

## 5 ANNEX

### 5.1 LITERATURE

Dickow, Marcel et al. (2015), *First Steps towards a Multidimensional Autonomy Risk Assessment (MARA) in Weapons Systems*, <[https://www.swp-berlin.org/fileadmin/contents/products/arbeitspapiere/FG03\\_WP05\\_2015\\_MARA.pdf](https://www.swp-berlin.org/fileadmin/contents/products/arbeitspapiere/FG03_WP05_2015_MARA.pdf)> (August 08, 2019).

Ekelhof, Merel (2018), *Lifting the Fog of Targeting: “Autonomous Weapons” and Human Control through the Lens of Military Targeting*, in: Naval War College Review: (71) No. 3, Article 6.

International Committee of the Red Cross, *Rule 15. Precautions in Attack*, IHL Database – Customary IHL, <[https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_rul\\_rule15](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule15)> (August 08, 2019).

International Committee of the Red Cross (1987), *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Commentary of 1987: Definition of attacks and scope of application*, <<https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=F5EA0CB6C1075C59C12563CD004345C3>> (August 08, 2019).

International Committee of the Red Cross (April 2018), *Statement at the CCW GGE on LAWS: Further consideration of the human element in the use of lethal force; aspects of human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems*, <[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/5216D20D2E98E7AAC12582720057E6FC/\\$file/2018\\_LAWS6b\\_ICRC1.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/5216D20D2E98E7AAC12582720057E6FC/$file/2018_LAWS6b_ICRC1.pdf)> (August 08, 2019).

International Panel on the Regulation of Autonomous Weapons (March 2018), *Focus on the Human-Machine Relation in LAWS*, <[https://www.ipraw.org/wp-content/uploads/2018/03/2018-03-29\\_iPRAW\\_Focus-On-Report-3.pdf](https://www.ipraw.org/wp-content/uploads/2018/03/2018-03-29_iPRAW_Focus-On-Report-3.pdf)> (August 08, 2019).

- International Panel on the Regulation of Autonomous Weapons (December 2018), *Concluding Report*, <[https://www.ipraw.org/wp-content/uploads/2018/12/2018-12-14\\_iPRAW\\_Concluding-Report.pdf](https://www.ipraw.org/wp-content/uploads/2018/12/2018-12-14_iPRAW_Concluding-Report.pdf)> (August 08, 2019).
- NATO (April 2016), *NATO Standard AJP-3.9. Allied Joint Doctrine for Joint Targeting*, <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/628215/20160505-nato\\_targeting\\_ajp\\_3\\_9.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/628215/20160505-nato_targeting_ajp_3_9.pdf)> (August 08, 2019)
- Roff, Heather & Moyes, Richard, (April 2016), *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons*, <<http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>> (August 08, 2019).
- Sauer, Frank et al. (2018), *Autonomy in Weapon Systems. The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy*, Heinrich Böll Foundation, <[https://www.boell.de/sites/default/files/boell\\_autonomy-in-weapon-systems\\_v04\\_kommentierbar\\_1.pdf?dimension1=division\\_oen](https://www.boell.de/sites/default/files/boell_autonomy-in-weapon-systems_v04_kommentierbar_1.pdf?dimension1=division_oen)> (August 08, 2019).
- Society of Automotive Engineers (September 2016), *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016\_201609)*, <[https://www.sae.org/standards/content/j3016\\_201609/](https://www.sae.org/standards/content/j3016_201609/)> (August 08, 2019).

## 5.2 MEMBERS OF IPRAW

<p><b>Peter Asaro</b> Professor <i>The New School</i> New York, USA</p>	<p><b>Denise Garcia</b> Professor <i>Northeastern University</i> Boston, USA</p>	<p><b>Shashank Reddy*</b> India</p>
<p><b>Vincent Boulanin</b> Researcher <i>Stockholm International Peace Research Institute</i> Stockholm, Sweden</p>	<p><b>Erin Hahn</b> Researcher <i>Johns Hopkins University Applied Physics Laboratory</i> Washington D.C., USA</p>	<p><b>Heigo Sato</b> Professor <i>Takushoku University</i> Tokyo, Japan</p>
<p><b>Thompson Chengeta</b> European Research Council Fellow on Drone Violence &amp; AI Ethics <i>University of Southampton</i> United Kingdom</p>	<p><b>Vadim Kozyulin</b> Researcher <i>PIR Center for Policy Studies</i> Moscow, Russia</p>	<p><b>Frank Sauer</b> Researcher <i>Universität der Bundeswehr</i> Munich, Germany</p>
<p><b>Anja Dahlmann</b> Researcher <i>Stiftung Wissenschaft und Politik</i> Berlin, Germany</p>	<p><b>Ian MacLeod</b> Researcher <i>Johns Hopkins University Applied Physics Laboratory</i> Washington D.C., USA</p>	<p><b>Lena Strauß</b> Research Assistant <i>Stiftung Wissenschaft und Politik</i> Berlin, Germany</p>
<p><b>Marcel Dickow</b> Head of Research Division <i>Stiftung Wissenschaft und Politik</i> Berlin, Germany</p>	<p><b>AJung Moon</b> Assistant Professor <i>McGill University</i> Montreal, Canada</p>	<p><b>Binxin Zhang</b> Assistant Professor <i>Xiamen University Law School</i> Xiamen, China</p>

The asterisk indicates those members, who did not participate in the 8<sup>th</sup> meeting of iPRAW (“Human Control, Verification, Proliferation”) in May 2019.

**Former members:** **Liran Antebi**, Research Fellow at the Institute for National Security Studies in Tel Aviv, Israel; **Deane-Peter Baker**, Senior Lecturer at the University of New South Wales Canberra in Australia; **Robin Geiß**, Professor at the University of Glasgow in Glasgow, United Kingdom; **Dong Lin**, Researcher at the National University of Defense Technology in Changsha, China; **David Hyunchul Shim**, Professor at the Korea Advanced Institute of Science and Technology in Daejeon, South Korea; **Kelvin Wong**, Unmanned Systems Editor at *Jane’s International Defence Review* in Singapore.

International Panel on the Regulation of Autonomous Weapons (IPRAW)

coordinated by:

Stiftung Wissenschaft und Politik (SWP) – German Institute for International and Security Affairs  
Ludwigkirchplatz 3-4  
10719 Berlin, Germany

August 2019

[www.ipraw.org](http://www.ipraw.org)  
[mail@ipraw.org](mailto:mail@ipraw.org)

This project is financially supported by the German Federal Foreign Office.

