



Concluding Report: Recommendations to the GGE

December 2018

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

December 2018

www.ipraw.org
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 2018.

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 Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) 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) and the Johns Hopkins University Applied Physics Laboratory (JHU APL) are jointly organizing 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 how LAWS should be defined as well as on suggesting possible approaches to regulation. The panel's working sessions will cover the following topics

- state of technology and operations as well as existing definitions of LAWS
- computational systems within the scope of LAWS
- autonomy and human control
- ethics, norms and public perception
- risks and opportunities
- IHL and other fields of law.

iPRAW will publish working documents on each of these topics and will, in addition, publish the panel's final recommendations aimed at informing the CCW process.

Procedure: The participants commit themselves to actively participate in and contribute to all meetings and the scientific dialogue in-between meetings. The panel will meet seven times over the course of two years, starting in March 2017. Each meeting will take two and a half days and will be hosted by SWP in Berlin. Papers with agreed upon recommendations on relevant issues will be drafted and published via the project's website (www.ipraw.org) in-between meetings.

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. Please direct your questions and remarks about the project to mail@ipraw.org.

CONTENT

- Executive Summary.....5
- 1 Introduction8
- 2 Human Control.....11
 - 2.1. Necessity for Human Control.....12
 - 2.2. iPRAW’s Concept of Human Control.....14
 - 2.3. Challenges to the Concept of Human Control.....15
- 3 Regulatory Options and their Implications for Operations, Law, and Ethics.....17
 - 3.1. A Principle of Human Control.....17
 - 3.2. Mapping of Regulative Options.....18
- 4 Conclusion22
 - 4.1. Recommendations22
 - 4.2. Remaining Questions23
- 5 Annex25
 - 5.1. Literature.....25
 - 5.2. Members of iPRAW27

FIGURES & TABLES

- Figure 1: United States of America’s Dynamic Targeting Cycle.....9

- Table 1: Requirements for Human Control over the Use of Force14
- Table 2: Regulatory Options of Global Governance to Implement the Principle of Human Control21

EXECUTIVE SUMMARY

Over the course of two years, iPRAW has identified and analyzed some of the most pressing issues related to the ongoing debate on LAWS within the CCW. In that time, the group developed and refined a series of recommendations gained from insights from the analysis. This report is the culmination of iPRAW's work to date, and captures the panel's recommendations to the CCW States Parties. In all but one instance, there is consensus within the group on each recommendation, and in the instance where there are differing views, they are explained as the logic of each approach may also be useful to the CCW. It is iPRAW's view that the CCW is at a pivotal point, where in particular, further work by states to elaborate on the concept of human control in weapon systems is critical to moving toward any type of binding or non-binding regulatory instrument. LAWS present challenging operational, legal, and ethical issues and iPRAW considers it important for states parties to take regulatory action, legally binding or not, as justified by analysis, to shape whether and how LAWS are developed.

Human Control as a Principle and Concept

In the current debate on LAWS, the principle – understood as a guiding norm to derive concrete behavior from – of human control is firmly rising as a prescriptive blueprint for action. The majority of CCW States Parties seems to agree that human control should constitute the basis of new norms regarding LAWS.

iPRAW recommends that a principle of human control should be internationally recognized within the CCW and possibly other documents of international law and be the basis from which requirements can be developed as part of a norm-shaping process. Based on this principle, iPRAW developed the following concept of human control to apply to the use of force: The design of weapon systems with autonomous functions must enable the operator to *understand* the operational context to allow for informed decisions over each step of the use of force. 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 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 mission, which is how these capabilities will be developed and fielded.

	Situational Understanding	Intervention
Control by Design (Technical Control)	Ability to monitor information about environment and system	Modes of operation that allow human intervention and require them in specific steps of the targeting cycle
Control in Use (Operational Control)	Appropriate monitoring of the system and the operational environment	Authority and accountability of human operators, teammates and commanders; abide by IHL

Requirements for Human Control over the Use of Force

Mapping of Regulative Options

CCW States Parties have several regulatory options to acknowledge and strengthen the principle of human control. Those can be legally or politically binding prohibitions or obligations in various forms. Factors besides the binding nature and force of an agreement are the substance of content and the inclusiveness of the document (which, in the CCW means consensus of all states parties).

iPRAW identifies three **legally binding** options, that would have similar implications but vary in their extend to security issues: (1) a positive obligation to safeguard human control in the use of force, (2) a comprehensive ban of the development and use of LAWS, (3) a ban of the use of LAWS. Those binding options could be prepared or accompanied by **soft law** like a political declaration. Such instruments could strengthen the principle of human control but would leave most details to the discretion of each individual state. Each of those options has certain operational, legal, ethical, and security implications. Depending on the perspective of states parties those carry different advantages and disadvantages.

If the CCW States Parties cannot find a consensus on the regulation of LAWS, national weapon reviews like those based on **Article 36** Additional Protocol (I) of the Geneva Conventions would constitute a fallback option. Since such weapon reviews are not a universal practice and require a fairly low threshold, iPRAW regards this legal instrument as necessary but insufficient as a stand-alone action.

Recommendations

Accordingly, the first consensus-based conclusion is that **inaction vis-à-vis the creation of global governance is not a viable option** regarding the issues raised by autonomy in weapon systems. All members of iPRAW recognize that LAWS raise fundamental and pressing questions and that the accompanying ethical, legal, and political implications are too wide-ranging and vital to remain unaddressed.

The second consensus-based conclusion is that **human control has to be the foundation of any policy formulated, be it legally or politically binding**. All iPRAW members agree that human control in human-machine interaction and machine dependence on humans in the execution of the targeting cycle must be retained. iPRAW stipulates that the control exercised by the human operator must be sufficient to reflect the operator's intention to establish the legal accountability and ethical responsibility for all ensuing acts. Furthermore, iPRAW suggests formulating **minimum requirements for human control** and offers its analytical framework of control by design and in use as well as the need for situational understanding and the ability for intervention as a guiding principle.

In addition to that, the panel formulated two divergent positions on the need for a legally binding regulation but agrees that the principle of human control should be formally recognized internationally and that states should implement respective standards to apply it in new acquisition programs.

In a second phase iPRAW will continue its work in 2019 to address remaining questions like details on the concept of human control, verification, and non-proliferation.

1 INTRODUCTION

The rise of lethal autonomous weapon systems (LAWS) enabled by advances in technology raises fundamental questions about maintaining human control from an operational, legal, and ethical perspective. This type of weapon system, which affords new capabilities to militaries able to develop them, changes the human-machine relationship in ways that require new thinking about the role and responsibilities of humans, and the mechanisms employed to effect human control in use including during combat. The goal of iPRAW is to understand these potential challenges and make analytically-backed recommendations to the Group of Governmental Experts as part of the Convention on Certain Conventional Weapons (CCW) based upon the diverse expertise from our members.¹

Framework for Analysis

iPRAW adapted the United States' dynamic targeting cycle as the foundation for a framework in which to evaluate the operational, technical, legal, and ethical issues within LAWS.

A targeting cycle lays out the functional steps and decisions involved with military targeting, modified versions are employed by many advanced military forces around the world. This approach helps to deconstruct the various LAWS issues in a structured way, which helps to identify how technology could impact lethal decision-making in each step of the targeting process. It should be noted that this targeting cycle does not stand alone and is informed by human decisions made as part of the larger military operation. The U.S. version of this decision-making process includes the six steps “find, fix, track, target, engage, and assess” which is often referred to by

¹ The *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 as amended on 21 December 2001 (CCW)*. It is referred to as the Convention on Certain Conventional Weapons and it is also known as the Inhumane Weapons Convention.

the acronym F2T2EA. It encompasses all major actions and decisions involved with lethal dynamic targeting decisions by U.S. forces and allies. Figure 1 describes the purpose, decisions, and current methods involved with the steps of F2T2EA.

iPRAW uses this framework in its analysis because it provides an understandable common framework between different entities and stakeholders in the LAWS discussion. The structure of the dynamic targeting cycle illuminates how many lethal decisions are carried out in language that is accessible to technologists, jurists, activists, and ethicists. iPRAW found that this structure provides greater clarity to the analysis, which allows to focus on points of concern in those subject areas, rather than an *ad hoc* debate on particular issues to the exclusion of other, potentially more relevant ones. **In particular, this approach showed how lethal decision-making is expansive, sequential, and cumulative, taking shape in each step of the F2t2EA process.** This is a notably different and more in-depth approach than the common focus on the targeting and engagement steps or more broadly applied *critical functions* of selection and engagement. Indeed, it helps illuminate that to focus solely on the final steps could overlook a series of autonomous actions that may occur earlier in the cycle but with potentially significant impact on the latter two steps, whether machine autonomy is applied there or not. This framework is adopted throughout iPRAW's analysis of different topics covered at individual meetings, e.g. computational methods within the scope of LAWS, autonomy and human control, or ethics, norms, and public perception, to name a few.

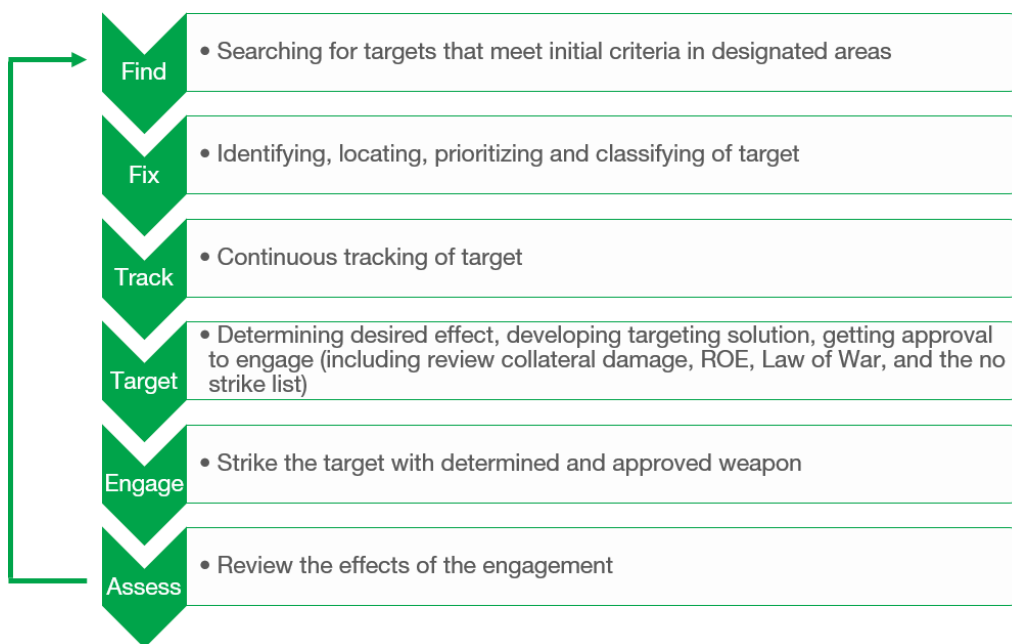


Figure 1: United States of America's Dynamic Targeting Cycle

Key Terms

A major challenge within the larger debate has been whether and how to define and operationalize the practical meaning of the term “autonomy” in lethal autonomous weapon systems. iPRAW came to use the term **machine autonomy** to mean specific functions performed by a machine without direct human control. This definition worked well with the targeting cycle framework and is the natural counterpart to direct *human control*. These two terms exist on the same conceptual sliding scale – more machine autonomy means less human control over the machine, and more human control necessarily translates into less machine autonomy.² This relationship underscores the importance of discussing these two terms jointly.

Further, iPRAW adopted the term **computational methods** to refer to the technology and techniques used to enable machine autonomy. The panel adopted this in place of artificial intelligence to avoid the bias that occurs when technology is described in anthropomorphized terms, which run the risk of evoking unrealistic notions of capabilities or unnecessary philosophical debates.

² For the difference between direct manipulation and control see iPRAW (2018a), *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> (December 12, 2018), pp.12-13.

2 HUMAN CONTROL

While much attention has centered on how to define autonomy, this focus did not bring the debate closer to clarification on why the definition matters or how it impacts whether to pursue regulation. In the course of iPRAW's analysis, the limitations of this definitional focus 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 definitional debate 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 environment and human machine relationship and how they may be impacted with differing applications of machine autonomy. In previous reports, iPRAW proposed a concept for minimum requirements of human control.³ Building on that, states can consider probing various possible implementations of control and how they impact the human-machine relation.

The international debate about LAWS references a number of different concepts of this relation already. They go under names such as human involvement,⁴ appropriate levels of human judgment⁵ or, most prominently, meaningful human control⁶. iPRAW acknowledges these concepts as valuable contributions to the debate and adopts a complementary perspective on the term *control* to enhance the current evolving understanding of human control and, in addition, to facilitate an interpretation of operational consequences for the use of weapon systems with autonomous functions.

³ See iPRAW (2018a), pp. 14-17.

⁴ See CCW (2016), *Recommendations to the 2016 Review Conference – Advanced Version, Submitted by the Chairperson of the Informal Meeting of Experts*.

⁵ See e.g. United States Department of Defense, *Directive 3000.09*, p. 2.

⁶ See Heather Roff & Richard Moyes (2016), *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons*, Article 36, <<http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>> (December 12, 2018).

2.1. NECESSITY FOR HUMAN CONTROL

Human control is a keystone for ethical concepts like human dignity, legal concepts of agency, attribution, and accountability, state authority and command responsibility, as well as military operational control of forces to achieve ends within political values. Technology has advanced, but direct and indirect control by humans has endured as a necessity, as technology cannot provide the capability needed for LAWS to be used in compliance with all legal obligations and operational requirements. The advance of technology that enables LAWS may change the degree, and type of control that exists within the bounds of operational, legal, and ethical acceptability, but at this point in time, will not eliminate it. Furthermore, it is possible it may never eliminate the need for human control depending on how states interpret ethical arguments and legal obligations. The question of whether it can or should be eliminated is at the core of the CCW debate. Statements made by states parties indicate human control as a principle must persist, even if there is disagreement on what human control means and the degree to which it can be delegated.

In some instances, the application of **computational methods** in the targeting cycle can lead to better outcomes than human performance alone could. Nevertheless, our discussion showed the limitations of computational methods as enabling technologies in the military domain and highlighted that they most likely cannot replace the unique judgment of the human decision-makers. Any complex computational system consists of modular subsystems, each of which inherently has limitations and points of failure. Applying multiple computational systems across each step of the targeting cycle may result in cumulative failures that can be catastrophic and hard to anticipate. Any system that executes sequential processes, such as selecting and engaging targets, can be subject to path dependencies where errors or decisions, in any step, can propagate and reverberate throughout the rest of the sequence, i.e. steps of F2T2EA.

iPRAW uses the term **computational methods** to address mathematical techniques often referred to as **artificial intelligence (AI)** and **machine learning**. When mentioning algorithms that sense, recognize, plan, decide, or act autonomously, we do not mean to anthropomorphize machines. Instead, these terms should be understood as shorthand descriptions.

Militaries utilize a range of mechanisms to achieve and maintain control over the use of force to operate within national values and international law, to achieve operational ends effectively, and prevent fratricide, which are all based on precision, predictability, and lethal efficiency. The application of operational art along with weighing of military necessity, risk, and priorities all rely on the judgment of human commanders. This kind of control is exercised through military decision-making processes that include mechanisms such as targeting cycles.

From a **legal perspective**, human control is relevant to maintain accountability and comply with core rules of international humanitarian law (IHL), especially the requirement of proportionality and military necessity.⁷ At this point in time, it is unclear whether autonomous systems will be able to comply with these rules:

⁷ The principle of distinction between civilians and combatants also requires human control. In this case, technical solutions are easier to find, though, as it (mostly) calls for a less complex situational understanding than proportionality and necessity assessments.

proportionality and military necessity are value judgment determinations that take into account factors that are not easily modelled but are bound by IHL parameters.⁸ The International Committee of the Red Cross (ICRC) argues that these legal judgments always require a human decision maker to apply the law to the specific situation at hand.⁹ Current interpretations by some states are similar.

In addition to proportionality and necessity there is precaution, a foundational and guiding applicable rule of IHL: the precautionary principle requires that actors take actions to prevent harm. Human control as a way to enable compliance with the precautionary principle is essential. It is possible that technology could enable humans to more readily comply with this principle, but it is unclear whether it could ever develop to the point where it could comply on its own.

The concept of **human dignity**, linked to IHL through the Martens Clause and customary law, is seen by some as adding to the requirement for human control. Here, human control safeguards moral agency to comply with ethical requirements. Human dignity has been articulated in various world philosophies and religions, has been integral to legal theory and practice, and has been codified in the constitutions of most countries as well as the Preamble of the United Nations Universal Declaration of Human Rights. In addition, the almost universally adhered to and legally binding International Covenant on Civil and Political Rights of 1966 codified the necessity of upholding human dignity.¹⁰ This provides strong support that human dignity has both a defined structure and is broadly recognized as integral to law and morality. iPRAW analyzed a set of minimum requirements for human dignity in the use of force as the ability to:¹¹

1. **Recognize** a human being as a human, not just distinguish it from other types of objects and things but as a being with rights that deserve respect;
2. Understand the **value of life** and the significance of its loss; and
3. **Reflect** upon the reasons for taking life and reach a rational conclusion that killing is justified in a particular situation.

Depending on the moral position, one would assume or deny that autonomous functions in weapon systems break the link to moral agency. In the first case, it would be necessary to safeguard moral agency through human control, in the latter case

⁸ See Alan Schuller (2017), “At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law”, in: *Harvard National Security Law Journal*, Vol. 8, pp. 379-425.

⁹ See International Committee of the Red Cross (2018a), *Towards Limits on Autonomy in Weapon Systems*, Statement, <<https://www.icrc.org/en/document/towards-limits-autonomous-weapons>> (December 12, 2018); International Committee of the Red Cross (2018b), *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*, Statement, <[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/5216D20D2E98E7AAC12582720057E6FC/\\$file/2018_LAWS6b_ICRC1.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/5216D20D2E98E7AAC12582720057E6FC/$file/2018_LAWS6b_ICRC1.pdf)> (December 12, 2018).

¹⁰ Article 10: All persons deprived of their liberty shall be treated with humanity and with respect for the inherent dignity of the human person.

¹¹ See Peter Asaro (2016), “Jus Nascendi, Robotic Weapons and the Martens Clause”, in: Ryan Calo, Michael Fromkin & Ian Kerr (eds.), *Robot Law*, Edward Elgar Publishing, pp. 367–386.

one would want to safeguard the ability to use a weapon system lawfully at the current state of technology. In consequence, both positions would require human control in both the design of the system and its use. **Inherent in both views is an acknowledgment – tacit or explicit – of the principle of human control.**¹²

Therefore, iPRAW recommends that the principle of human control should be internationally recognized within the CCW and possibly other documents of international law¹³ and be the basis from which requirements can be developed as part of a norm-shaping process. What this principle means in practice will need to emerge in the CCW from the requirements states develop, formally internationally agreed upon and ideally also through domestic processes. Weapon systems should be developed with this principle in mind.

iPRAW notes that the critical question emerging from these findings concerns where in the process the link to human control is interrupted or broken. This assessment is at the heart of the debate and should be the focus of ongoing CCW discussions. As a starting point, iPRAW suggest the following concept of human control.

2.2. IPRAW’S CONCEPT OF HUMAN CONTROL

The design of weapon systems with autonomous functions must enable the operator to *understand* the operational context to allow for informed decisions over each step of the use of force. 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 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 mission, which is how these capabilities will be developed and fielded.

	Situational Understanding	Intervention
Control by Design (Technical Control)	Ability to monitor information about environment and system	Modes of operation that allow human intervention and require them in specific steps of the targeting cycle
Control in Use (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 over the Use of Force

¹² See iPRAW (2018b), *Focus on Ethical Implications for a Regulation of LAWS*, <https://www.ipraw.org/wp-content/uploads/2018/08/2018-08-17_iPRAW_Focus-On-Report-4.pdf> (December 12, 2018), pp.15-18.

¹³ For a suggestion on one possible path of implementation see Elvira Rosert (2017), *How to Regulate Autonomous Weapons. Steps to Codify Meaningful Humanitarian Control as a Principle of International Humanitarian Law*, PRIF Spotlight 6/2017, Frankfurt/M.

2.3. CHALLENGES TO THE CONCEPT OF HUMAN CONTROL

With regard to the presented concept of human control, iPRAW identified a set of challenges and open questions that call for further research and discussion in the CCW framework. The most relevant ones are (1) the influence of the context of use on the necessary level of human control, (2) the relation between a weapon system's life cycle and the targeting cycle, and (3) the potential benefits of autonomous targeting functions in a restricted operational application ('boxed autonomy').

Human control, as argued in previous iPRAW 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, especially in situations with dynamically changing contexts. iPRAW suggests to further study the influence of specific variables, such as the domain of the operation (air, land, sea, space)¹⁴ and the frequency of updates to ensure situational understanding, on the implementation of control in use and the necessity of the operator/commander for intervention. To this end, it is of paramount importance to define minimum requirements for human control. This should include points of time of human decision making in the targeting cycle. While the panel could not agree on an adequate description of an ultimate point of time, we can state that human decision making is a necessity and should take place as close as possible to the engagement (i.e. release of the weapon or use of force) – having in mind, that a weapon system can be programmed, tasked and activated significantly earlier than the actual engagement takes place.

iPRAW's approach to human control is focused on the necessity of humans 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 e.g. research and development, programming and testing of systems, and components already provide for the application of human control in operation. While responsible innovation and research is a key element to shape 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 human control for the development of new weapon systems. This particularly includes the commercial sector.

iPRAW's discussions show that 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. This may lead to higher precision of weapon systems. Furthermore, iPRAW acknowledges that **operational demands** may incentivize restrictions on the communication to and from a weapon system during mission. Loss of communication due to

The concept of 'boxed autonomy' consists of a predefined context in which the system has to locate and engage a target. The box conditions would be preprogrammed and combined with parameters limiting the system's abilities once it is outside of the range of communication.

¹⁴ For the time being, iPRAW excludes the cyber domain from its analysis and focuses on tangible systems. It should be noted, though, that LAWS and autonomous cyber weapons share a set of challenges with regard to international humanitarian law.

technical error or adversary interference may also lead to a loss of human control during operation while it might be of military interest to continue the mission autonomously. iPRAW has been skeptical about these scenarios of boxed autonomy and therefore urges **states to further investigate and exchange their positions regarding these specific situations.** We emphasize that a potential regulation should take into account minimum requirements for human control to ensure that humans remain in control over the weapon systems they use.

3 REGULATORY OPTIONS AND THEIR IMPLICATIONS FOR OPERATIONS, LAW, AND ETHICS

To acknowledge the principle of human control over the use of force, the CCW States parties have several general options, some of which are not mutually exclusive and could build on each other. They include a legally binding protocol prohibiting LAWS (and in turn requiring states to safeguard human control in the use of force), legally binding restrictions of the use of LAWS, a political declaration, a code of conduct or best practices, to name a few. States parties could also conclude that existing state mechanisms to safeguard human control, e.g. Article 36 weapon reviews are sufficient. In addition to or in lieu of that, states parties may find multilateral agreements outside of the CCW framework comparable to the bans of anti-personnel mines and cluster munitions.

3.1. A PRINCIPLE OF HUMAN CONTROL

Universal principles, such as precaution in war and the distinction between civilians and combatants give rise to new norms of behavior that guide the conduct of states at war and peace. These have been codified under IHL. iPRAW understands the term *principle* as a guiding norm that other, more concrete norms (and subsequent concepts) can derive from. Principles express values and judgments, the respective norms would be behavioral rules.¹⁵

In the current debate on LAWS, the principle of human control is firmly rising as a prescriptive blueprint for action. The majority of CCW States Parties seems to agree that human control is the principle that should constitute the basis of new norms of behavior regarding LAWS.¹⁶ Additionally, the norm of human control finds at least implicit acknowledgement and possible antecedent in various existing arms control

¹⁵ Based on a presentation by Prof. Elvira Rosert during iPRAW's sixth meeting in June 2018.

¹⁶ At the very least, the notion that human involvement in the use of force is necessary has not been contested during the CCW deliberations.

treaties. Human Rights Watch identifies the norm in the bans on biological and chemical weapons as well as anti-personnel landmines.¹⁷ **iPRAW recommends that a principle of human control should be internationally recognized within the CCW and possibly other documents of international law and be the basis from which requirements can be developed as part of a norm-shaping process.** Taking this requirement of human control as a departing point and a basis for international action, several regulatory options arise. iPRAW recommends states to take them into account.

3.2. MAPPING OF REGULATIVE OPTIONS

As mentioned above, CCW States Parties have several regulatory options to acknowledge and strengthen the principle of human control (for an overview see Table 2 on page 21). Those can be legally or politically binding prohibitions or obligations in various forms.¹⁸ Factors besides the binding nature and force of an agreement are the substance of content and the inclusiveness of the document (which, in the CCW means consensus by all states parties). The following options range from legally binding (hard law) options to soft law measures with varying substance of content.

Each of those options has certain operational, legal, ethical, and security implications. Depending on the perspective of states parties those carry different advantages and disadvantages. For example, the focus on human control in the design and use would have certain operational implications. Human oversight and options for intervention would call for a fighting speed adequate for human reaction time. On the one hand, this can be an operational disadvantage, especially if confronted with adversary LAWS. On the other hand, the comparably slower pace and human understanding might mitigate the risk of conflict escalation as described in Altmann/Sauer 2017.¹⁹

Hard Law

One option is a **legally binding protocol** that sets a **positive obligation** to safeguard human control in the use of force. That could encompass certain rules for the use of weapon systems or address requirements of the design phase. Such a broad regulation would not be tied to a certain weapon system, but would apply to (conventional) weapons in general and might even create a customary norm beyond the scope of the CCW. The protocol might include incentives like information sharing or technology transfer to motivate states to follow the obligation set by the new instrument. The details of implementing human control in the use of force could also

¹⁷ See Human Rights Watch (2016), *Killer Robots and the Concept of Meaningful Human Control. Memorandum to Convention on Conventional Weapons (CCW) Delegates*, <https://www.hrw.org/sites/default/files/supporting_resources/robots_meaningful_human_control_final.pdf> (November 29, 2018), pp. 10-12.

¹⁸ For an in-depth analysis on the future of global regulatory options also see: Denise Garcia (2018), "Lethal Artificial Intelligence and Change: The Future of International Peace and Security", in: *International Studies Review* 20, pp. 334-341.

¹⁹ See Jürgen Altmann & Frank Sauer (2017), *Autonomous Weapon Systems and Strategic Stability*, in: *Survival* 59 (5), pp. 117-142.

be left to an instrument like a code of conduct, which would be of a soft law nature nevertheless.

A positive obligation of human control over the use of force would be less vulnerable to technological advances than a technical definition of LAWS. It would address substantial ethical and legal concerns, like the violation of human dignity and the necessity of a human to make legal judgments – those concerns are not rooted in technological capabilities but are inherent in ethics and international law that set restraints on the waiving of human decision making to machines.²⁰

Another option is a protocol for a **comprehensive ban on the development and use of LAWS**. With regard to the principle of human control, LAWS could be defined as systems that lack human control (in the targeting functions). A more technical definition based on certain capabilities (e.g. understanding of intent, certain machine learning techniques) might be less future proof as the technology develops rapidly and outcomes are hard to predict.

Nevertheless, a comprehensive ban could encompass certain generalized design characteristics like mandatory communication links. An annex listing explicit exclusions of existing systems designed to target incoming munitions might be necessary; a regular review of this annex to match recent technology developments could be prudent.

Overall, the implications of a comprehensive ban would resemble those of a positive obligation. Depending on the specific definition with regard to human control the ban could be more selective than the positive obligation. It would therefore lead to a smaller impact on international arms control in general because it would be tied to a narrower spectrum of weapon systems.

A **protocol that solely bans the use** of LAWS, as opposed to ban the development and use, would allow states to continue to research and develop such systems. Similar to that, a CCW protocol could set limitations to **certain uses** of LAWS short of a ban. Such a specific legally binding restriction could be a context-dependent prohibition that focusses on specific target classes, employment parameters like range or time of application, or types of operational environments (i.e. boxed autonomy).

All three legally binding options presented above would support legal accountability and safeguard human dignity through moral agency. While a comprehensive ban might also address certain security concerns by limiting proliferation and arms dynamics,²¹ the mere ban of use would not include this aspect as it allows development, domestic use, and export of LAWS technologies. Those elements are not necessarily the subject of CCW deliberations, they are a dimension of LAWS nonetheless.

²⁰ See iPRAW (2017), *Focus on Computational Methods in the Context of LAWS*, <https://www.ipraw.org/wp-content/uploads/2017/11/2017-11-10_iPRAW_Focus-On-Report-2.pdf> (December 12, 2018), p. 15-18; iPRAW (2018b), p. 13; also ICRC (2018a) and (2018b).

²¹ See Altmann & Sauer (2017).

Soft Law

There are multiple regulatory options short of a legally binding regulation that the CCW could consider. Such soft law instruments could entail best practices, a code of conduct or legal interpretation guidance similar to the Tallinn Manual²². These options could encompass transparency measures or guidelines for manufacturers and R&D, and could offer guidance for practices of weapon reviews. They would only have hortatory force, but could set prescriptions and have a harmonizing effect on national legislation or pave the way for further building of new or expanded regulatory options. As soft law is not legally binding, it does not necessarily require explicit enforcement or verification mechanisms that might be part of the binding options presented above.

In this context, a **political declaration** of the CCW high contracting parties would be another way to acknowledge and strengthen the principle of human control. Such a declaration would set initial behavioral constraints on many states as it creates a political obligation that cannot be ignored. It would be a guidance for national legislation, leaving the details to the discretion of each individual state. The establishment of a technology committee could enhance the scope by pointing at certain issues (similar to the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) or the Organisation for the Prohibition of Chemical Weapons (OPCW)).

Other forms of **regulation below a legally binding document** could still create and strengthen the principle of human control, being potentially a first step towards further regulation. The implications for the establishment of a subsequent legally binding document are unclear, though: those political measures could become a stepping stone for more formal regulation, but they might as well represent the end of discussions for some states.

No Regulation

Apart from those actions, another outcome of the CCW process could be **inaction**. That could result from lack of consensus about any regulatory approach or the perception by all or some states parties that existing IHL and national weapon reviews are sufficient to address the challenges related to autonomous functions in weapon systems. In this case, the most prominent legal restriction would be national weapon reviews, especially Article 36 Additional Protocol (I) to the Geneva Conventions²³.

Article 36 imposes an obligation on states to ensure the use of weapons, means or methods of warfare is lawful prior to making new systems operational. It is often referred to as a ‘weapon review’ or ‘Article 36 review’, and it is left to states to determine their own domestic processes for evaluation. Article 36 is necessary but not sufficient: First of all, is not applied universally, as very few states have such a process in place. Second, it requires only a determination that weapons do not violate IHL (and possibly international human rights law) in general, a fairly low

²² Tallinn Manual on the International Law Applicable to Cyber Warfare, 2013.

²³ Full title: 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.

threshold to meet since just one IHL conform application is sufficient. Given the increasing innovation of weapons systems, it may become more and more difficult for a commander to understand how a system works and evaluate whether it will be lawful to use it in a given situation absent a supplemental review or process.

Moreover, the testing and evaluation of systems with computational methods is costly and presents several other challenges, which may translate to reviews which include incomplete information or cannot quantify the reliability of the system. Nonetheless, Article 36 reviews remain important, and with additional processes or guidance such as that recommended by Boulanin/Verbruggen (2017),²⁴ could make the review more robust and increase the likelihood of compliance with international law. The challenge however is to universalize the practice of weapons review and make it more transparent.

Regulatory Option	Explanation
Hard Law	
Positive Obligation	A proscriptive legal obligation that binds state parties to its terms, the scope could be broad or only apply to aspects of LAWS
Comprehensive Ban	A binding legal instrument to development, acquisition, and use of LAWS
Ban of Use	A binding legal instrument that applies only to use of LAWS, not development or acquisition
Soft Law	
Political Declaration	A statement of agreed-upon standards but which is not legally binding
Others (e.g. best practices, code of conduct)	This category includes instruments such as guidelines for appropriate development and use or implementation of the technology, i.e. non-binding guidance, best practices, and potentially domestic law developed and adopted by a state
Inaction	
Solely Article 36	This approach relies on existing international law which requires states to review weapons to ensure they can be used in accordance with IHL
iPRAW analyzed the implications of those regulatory options along in their operational, legal, ethical, and (to some extend) security dimension.	

Table 2: Regulatory Options of Global Governance to Implement the Principle of Human Control

²⁴ See Vincent Boulanin & Maaike Verbruggen (2017), *Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies*, SIPRI, <<https://www.sipri.org/publications/2017/other-publications/article-36-reviews-dealing-challenges-posed-emerging-technologies>> (December 12, 2018), pp. 22-25.

4 CONCLUSION

After two years of work, iPRAW uses its fifth report to present the consensus reached among members with regard to overarching conclusions and also presents differing views among panelists with regard to preferable policy options. Additionally, the work of the panel points towards future analyses and questions, which are presented at the end of this section, preparing the second phase of iPRAW's work.

4.1.RECOMMENDATIONS

Consensus

The first consensus-based conclusion is that inaction vis-à-vis the creation of global governance is not a viable option with regard to the issues raised by autonomy in weapon systems. All members of iPRAW recognize that LAWS raise fundamental questions and that the accompanying ethical, legal, and political implications are too wide-ranging and important to remain unaddressed. **iPRAW considers it important for States Parties to take regulatory action, as justified by analysis, to shape whether and how LAWS are developed.**

The second consensus-based conclusion is that human control has to be the foundation of any policy that is formulated. All iPRAW members agree that human control in human-machine interaction and machine dependence on humans in the execution of the targeting cycle must be retained. iPRAW stipulates that the control exercised by the human operator must be sufficient to reflect the operator's intention for the purpose of establishing the legal accountability and ethical responsibility for all ensuing acts.

Furthermore, iPRAW suggests formulating minimum requirements for human control and offers its analytical framework of control by design and in use as well as the need for situational understanding and the ability for intervention as a guideline.

Differing Positions

There was no consensus amongst panelists on which of the regulatory options discussed in this report to adapt as a final recommendation. iPRAW notes that some of the policies detailed in section 3 are not mutually exclusive. It is conceivable, for instance, that a political declaration could be followed, eventually, by a binding regulation.

Some panelists argued that, given the potentially far-reaching consequences of the technology, a policy option that is binding should be recommended – specifically, that there must be a codification of both the positive and negative obligations of states as far as the application of autonomy in weapon systems is concerned.

Other panelists argued that, at this point in time, a non-binding instrument should be recommended, with human control as the anchoring principle for future requirements and actions by states. Here, the reasoning is that while there are indeed concerns whether computational methods can meet legal standards and ethical principles, a binding regulation may preempt the development of beneficial applications of autonomy in weapons systems. These panelists acknowledge that additional analysis could change the formulation of this position, but advocate for an incremental approach that starts with the principle of human control, with definition stemming from requirements or guidelines developed by states internationally and domestically as a mechanism to shape norms for LAWS. This approach has the potential to build common ground and understanding from those who will develop and use the systems, leading to a stronger norm. With this approach even a legally binding comprehensive regulation remains a possibility: As advancements are made, it is entirely possible that states, with input from the technical and operational communities, will determine a threshold has been met, development beyond which should not be permitted.

While this position advocates for an incremental approach to norm shaping and LAWS, it is in full alignment with the other viewpoints that the principle of human control should be formally recognized internationally and that states should implement standards guaranteeing it in new acquisition programs.

4.2. REMAINING QUESTIONS

Although this report concludes the first chapter of iPRAW's work, it will continue in a second chapter in 2019. The panel identified a set of open questions and subjects that could be of relevance to the work of the CCW Group of Governmental Experts on LAWS and will guide iPRAW's upcoming deliberations.

(1) Details on human control: iPRAW presented its concept of human control as situational understanding and options for intervention in the design and use of the systems. Those categories are a first step to understand the necessary elements of human control in the use of force, but more work needs to be done to understand the operational implications and to develop a common understanding of the term.²⁵

²⁵ For valuable ideas about distributed human control across the targeting cycle see Merel Ekelhof (2018), *Autonomous Weapons: Operationalizing Meaningful Human Control*,

(2) Verification and validation: The legally binding options mentioned above could be complemented by a verification mechanism. The actual scope and technical implementation of verification would depend on the specific subject of regulation, but it could be worthwhile to look for common denominators and limitations. In addition to that, the information sharing regarding a technical validation of human control in the targeting cycle could be a helpful instrument for both binding and non-binding measures.

(3) Non-Proliferation: The question of proliferation is a topic beyond the scope of the CCW and could be discussed under export control regimes like the Wassenaar Arrangement²⁶. Nevertheless, a CCW regulation restricting the development and use of autonomous targeting functions might be a first step to address this challenge.

<http://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/> (December 12, 2018).

²⁶ The *Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies* of 1996.

5 ANNEX

5.1. LITERATURE

Altmann, Jürgen; Sauer, Frank (2017), *Autonomous Weapon Systems and Strategic Stability*, in: *Survival* 59 (5), pp. 117-142.

Asaro, Peter (2016), “Jus Nascendi, Robotic Weapons and the Martens Clause”, in: Ryan Calo, Michael Froomkin & Ian Kerr (eds.), *Robot Law*, Edward Elgar Publishing, pp. 367–386.

Boulanin, Vincent; Verbruggen, Maaïke (2017), *Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies*, SIPRI, <<https://www.sipri.org/publications/2017/other-publications/article-36-reviews-dealing-challenges-posed-emerging-technologies>> (December 12, 2018).

CCW (2016), *Recommendations to the 2016 Review Conference – Advanced Version, Submitted by the Chairperson of the Informal Meeting of Experts*.

Ekelhof, Merel (2018), *Autonomous Weapons: Operationalizing Meaningful Human Control*, <<http://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/>> (December 12, 2018).

Garcia, Denise (2018), “Lethal Artificial Intelligence and Change: The Future of International Peace and Security”, in: *International Studies Review* 20, pp. 334–341.

Human Rights Watch (2016), *Killer Robots and the Concept of Meaningful Human Control. Memorandum to Convention on Conventional Weapons (CCW) Delegates*, <https://www.hrw.org/sites/default/files/supporting_resources/robots_meaningful_human_control_final.pdf> (December 12, 2018).

International Committee of the Red Cross (2018a), *Towards Limits on Autonomy in Weapon Systems*, Statement, <<https://www.icrc.org/en/document/towards-limits-autonomous-weapons>> (December 12, 2018).

- International Committee of the Red Cross (2018b), *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*, Statement, <[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/5216D20D2E98E7AAC12582720057E6FC/\\$file/2018_LAWS6b_ICRC1.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/5216D20D2E98E7AAC12582720057E6FC/$file/2018_LAWS6b_ICRC1.pdf)> (December 12, 2018).
- iPRAW (2017), *Focus on Computational Methods in the Context of LAWS*, <https://www.ipraw.org/wp-content/uploads/2017/11/2017-11-10_iPRAW_Focus-On-Report-2.pdf> (December 12, 2018).
- iPRAW (2018a), *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> (December 12, 2018).
- iPRAW (2018b), *Focus on Ethical Implications for a Regulation of LAWS*, <https://www.ipraw.org/wp-content/uploads/2018/08/2018-08-17_iPRAW_Focus-On-Report-4.pdf> (December 12, 2018).
- Roff, Heather; Moyes, Richard (2016), *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons*, Article 36, <<http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>> (December 12, 2018).
- Rosert, Elvira (2017), *How to Regulate Autonomous Weapons. Steps to Codify Meaningful Humanitarian Control as a Principle of International Humanitarian Law*, PRIF Spotlight 6/2017, Frankfurt/M.
- Schuller, Alan (2017), “At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law”, in: *Harvard National Security Law Journal*, Vol. 8, pp. 379-425.
- United States Department of Defense (2009/ 2017), *Directive 3000.09*.

5.2. MEMBERS OF IPRAW

Liran Antebi
Research Fellow
Institute for National Security Studies
Tel Aviv, Israel

Peter Asaro
Professor
The New School
New York, USA

Deane-Peter Baker*
Senior Lecturer
University of New South Wales
Canberra, Australia

Vincent Boulanin
Researcher
Stockholm International Peace Research Institute
Stockholm, Sweden

Thompson Chengeta
Fellow
South African Research Chair in International Law, University of Johannesburg
Johannesburg, South Africa

Anja Dahlmann
Researcher
German Institute for International and Security Affairs
Berlin, Germany

Marcel Dickow
Head of Research Division
German Institute for International and Security Affairs
Berlin, Germany

Denise Garcia
Professor
Northeastern University
Boston, USA

Robin Geiß*
Professor
University of Glasgow
Berlin, Germany

Erin Hahn
Researcher
Johns Hopkins University Applied Physics Laboratory
Washington D.C., USA

Vadim Kozyulin
Researcher
PIR Center for Policy Studies
Moscow, Russia

Ian MacLeod
Researcher
Johns Hopkins University Applied Physics Laboratory
Washington D.C., USA

AJung Moon*
Senior Advisor for UN Secretary-General's High-level Panel on Digital Cooperation
USA

Shashank Reddy*
Researcher
Carnegie India
New Delhi, India

Heigo Sato
Professor
Takushoku University
Tokyo, Japan

Frank Sauer
Researcher
Universität der Bundeswehr
Munich, Germany

David Hyunchul Shim*
Professor
Korea Advanced Institute of Science and Technology
Daejeon, South Korea

Lena Strauß
Research Assistant
German Institute for International and Security Affairs
Berlin, Germany

The asterisk indicates those members, who did not participate in the seventh meeting of iPRAW in October 2018. Former members: **Dong Lin**, Researcher at the National University of Defense Technology in Changsha, China and **Kelvin Wong**, Researcher at IHS Janes in Singapore.

The panel has been supported by contributions by Julia Buchholtz, Frank Flemisch (FKIE), Jürgen Altmann (TU Dortmund), and Elvira Rosert (IFSH, University of Hamburg).

