Weaponized Artificial Intelligence & Stagnation in the CCW: A North-South Divide Written by Alena Zafonte

This PDF is auto-generated for reference only. As such, it may contain some conversion errors and/or missing information. For all formal use please refer to the official version on the website, as linked below.

Weaponized Artificial Intelligence & Stagnation in the CCW: A North-South Divide

https://www.e-ir.info/2018/11/01/weaponized-artificial-intelligence-stagnation-in-the-ccw-a-north-south-divide/

ALENA ZAFONTE, NOV 1 2018

One's position on a given policy "ultimately depends on what one is concerned about," a fact that calls attention to a fundamental aspect of the discussion of Lethal Autonomous Weapons Systems (LAWS) in the United Nations; given historical North-South discrepancies in the burdens of technology and present day disparities in the technological development of LAWS, the Global North-South divide is a conspicuous component of the discussions in the Convention on Certain Conventional Weapons (CCW) (UNIDIR 2017: 1). The inactivity in the CCW can be attributed to a variety of factors as states continue to struggle for any form of consensus (UNIDIR 2017; Private Meeting 1 2018). This paper will therefore seek to illuminate the extent to which North-South discrepancies in the technological development of LAWS have contributed to stagnation in the CCW by examining the ways in which North and South approaches to these uniquely difficult conversations on definitions and legal attribution differ. Given that the Global South is often disproportionately impacted by new weapons technologies and that weaponized AI is primarily being developed in the North it is only natural that these perspectives fall into relatively defined, diverging categories as outlined in this paper; said divergence likely contributes to stagnation in the CCW to a notable degree (Private Meeting 1 2018).

Framework Justification

North-South Divide

While some may argue that considering an issue from a Global North-South perspective has the potential to oversimplify the discussion, the contrary is true for the purposes of this paper for three reasons. First, given that the Non Aligned Movement (NAM), which consists mainly of Global South states, is among the chief champions banning the development of LAWS it is natural that the topic be examined from this lense (Garcia 2015; 'Profile' 2009). Furthermore, the majority of states which are developing LAWS are considered to be within the Global North (Marr 2018). Lastly, the term 'Global South' has evolved to replace 'Third World' or 'Developing World' with a more fluid, less hierarchical connotation and therefore allows for a more nuanced analysis rather than restricting the conversation to developed vs developing states (Wolvers et al). Within this framework, North-South terminology should therefore not be taken too literally, but instead understood in the wider context of globalization as many countries considered to be in the Global South are not, in fact, below the equator (Wolvers et al).

This globalized context is imperative to applying this framework effectively as it is this context which directs states' policy positions. Research has shown that a one percent reduction in violence would save spending equivalent to the total global investments in development— this implies that Global South states, which bear a more significant portion of the suffering perpetrated by emerging weapons technologies, therefore have a dual interest in regulating the development of LAWS; by limiting the development and use of such weapons, more funds would be available for development investment and less violence would occur (Private Meeting 1 2018). By many accounts there has been a breakdown of the structures for managing conflict and the 2018 Global Peace Index demonstrates a worldwide decline in peacefulness, making this context more important than ever (Institute for Economics and Peace 2018; Private Meeting 1 2018).

Written by Alena Zafonte

Technological Development and Terminology

Thomas Burri's conclusions tie into this North-South analysis quite well: in an examination of the politics of robot autonomy at the Darpa Robotics Competition, Burri identifies a crucial point: there is a "disjunction between the world of roboticists and the world of social scientists." (Burri 2016: 342) Burri explains that one side, the 'operators', is composed of individuals such as technicians, engineers and computer scientists who are directly involved in the development and functioning of these technologies (Burri 2016). Alternatively, political scientists, lawyers, and related actors are termed 'observers' as they do not directly interact with the technology (Burri 2016). This logic can also be applied to states as, much like Burri's operators are able to discuss technicalities more effectively, those states pursuing the development of weaponized AI inherently have greater capacity for institutional knowledge and will therefore address the concerns regarding killer robots from a different perspective (Private Meeting 1 2018). This paper will thus adapt Burri's framework in order to intelligently discuss the impacts of North-South discrepancies in technological development on the CCW; those states which are not seeking to produce weaponized AI will be termed 'Observers' and those which are advancing such technology will be referred to as 'Developers'. Developers are primarily in the Global North while Observers are largely in the South: this overarching framework is utilized in order to capture those states which may not fit neatly within their North/South partners by allowing them to be discussed from a capacity perspective while also maintaining the connection to their region (Garcia 2015; Private Meeting 2 2018).

Within the CCW, there is a clear division between these Observers and Developers— in other words, between those states which favor politically binding limitations on LAWS and those which do not (Garcia 2015). Though the first category may be subdivided into states which favor human control and those which favor a preemptive total ban, these two groups are aggregated as favoring regulation for the auspices of this discussion (Garcia 2015). Though the ranks of the Observers include Global North states such as France and Germany, this group is more largely constituted by the Non Aligned Movement which is composed primarily of Southern states (Private Meeting 2 2018; Garcia 2015; 'Profile' 2009). The other side is led by the United States (US) and Russia which continue to perpetuate the discussion of legal definitions in hopes of elongating talks to avoid the development of regulations; these states maintain that periodic 'weapons reviews' of emerging LAWS will be sufficient for upholding international laws and norms (Garcia 2015; UNIDIR 2017).

As seen throughout history, the Global South tends to disproportionately suffer the consequences of technological advances in weapons: anti-personnel landmines, cluster munitions, drones and others have all been employed primarily by Northern states while the Observers bear the most significant losses (Private Meeting 3 2018; Private Meeting 4 2018). While the development of autonomous weapons is by no means an exclusively Northern phenomenon, the majority of these advances are occurring outside the Global South. For example, the United States launched testing of an autonomous warship in 2016 which is anticipated to ultimately have notable offensive capabilities (Marr 2018; UNIDIR 2017). Similarly, the United Kingdom (UK) expects its fully-Al combat drone Taranis to be a viable replacement for human piloted drones in the Royal Air Force in the near future (Marr 2018). Many other Global North countries have similar aspirations in the field of weaponized AI, though some have yet to actively pursue their development (UNIDIR 2017). There are also Global South states which have begun developing these technologies— namely China and (debatably) South Korea; however, both of these states are often considered in the Global North as well, making them viable for Developer classification and further justifying this framework (Marr 2018; Wolvers et al).

Stagnition: Case Study Examples

In order to examine stagnation in the CCW from this perspective it is necessary to first understand the context in which discussions on autonomous weapons are occuring. As put by UNIDIR, the CCW is an arms control treaty with the purpose of banning or restricting the use of weapons which cause "unnecessary, unjustifiable or superfluous suffering to combatants or to affect civilians indiscriminately" (UNIDIR 2017: 4). Given this mandate, it can be argued that the CCW better caters to Global South concerns by design; however, given that Global North states are, on average, more powerful and are typically the Developers this does not often come to fruition (Private Meeting 2 2018). The CCW stagnation is therefore best considered by examining what may be two of the most important

Written by Alena Zafonte

components of the conversation: the issues of definition and attribution. Given that both of these topics relate to individual country capacity and state-specific security concerns this analysis will serve to advance the discussion of Developer vs Observer dissonance in approach. Each topic will be explained in brief before Developer-Observer analysis is provided.

Defining Autonomy

Definitions in the CCW are often technical in nature and refer to a minutely specific class of weapons (UNIDIR 2017; Human Rights Watch 2016). For example, the Mine Ban Treaty refers explicitly to anti personnel landmines, meaning that non-anti personnel landmines are excluded from its provisions (Human Rights Watch 2016; Private Meeting 2 2018). Given that weaponized artificial intelligence is an emerging technology and LAWS are still in development it is uniquely difficult to utilize this traditional approach— a technical approach is not easily applied to abstract, not-yetdeveloped technologies. According to the United Nations Institute for Disarmament Research (UNIDIR), there is dissonance in even the understandings of autonomy; though, as stated by UNIDIR, "autonomy is a characteristic, not a thing in and of itself" it is often not discussed as such due to the Developer-Observer variance outlined above (UNIDIR 2017: 19; Private Meetings 1 2018; Private Meeting 5 2018). Since a definition will ultimately demarcate what is regulated and what is not, such dissonance allows "considerable space for manoeuvre to adhere to the letter without adhering to the spirit of the definition," a fact that is only exacerbated by the pace at which this technology is advancing (UNIDIR 2017: 20). For example, states may determine that it is offensive autonomous technologies that must be banned or limited while defensive systems (for example, C-RAM or Israel's Iron Dome) may be allowed to maintain a degree of autonomy despite the fact that "an offensive and defensive system are physically identical one simply modifies the conditions under which it is permitted to engage." (UNIDIR 2017: 20) Even applications of autonomy that are not inherently weaponized such as reconnaissance or surveillance technology "could be rapidly weaponized either deliberately or in an ad hoc manner." (UNIDIR 2017: 20) The Developer-Observer divide is seen here as, in many instances, Observer states attempt to discuss unmanned aerial vehicles within the context of defining autonomy; Observer states have been largely impacted by the deployment of this technology, but fundamentally misunderstand automation and remote-control as autonomy (Private Meeting 5 2018).

This confusion is indicative of the the distinction between Observer and Developer states and highlights precisely why this divide is important to consider: there is an evident division in the capacity to understand and discuss LAWS as an emerging threat. Buri's operators view upholds a "technical understanding" of autonomy which requires a human operator and machine to work together "almost like in an organic symbiosis" to complete a task; this is akin to the stance of the majority of the states that would fall in the Developer category and aligns with how the CCW has traditionally addressed such issues (Burri 2016: 354; Human Rights Watch 2016). The Observers' view as explained by Burri is nearly identical to the position of our Observer states: "the Observers' view focuses on the big picture, on what is still to come in light of what exists now" and views autonomy as "something uncontrolled, not predetermined and unpredictable." (Burri 2016: 354) Indeed, while Observer states are primarily concerned with the uncertainty of what lies ahead in the development of increasingly autonomous weapons, there are many Observers which seem to misunderstand autonomy as a characteristic, instead focusing on automated or unmanned technology such as drones or automated defense systems as mentioned (UNIDIR 2017; Private Meeting 5 2018).

As noted, "it is natural that proponents and opponents of AWS will seek to establish a definition that serves their aims and interests;" however, it is only possible to do so with a sufficient understanding of the potential effects of these weapons, meaning that the fast pace of development has impeded the effectiveness of these discussions and put the Observer states (read: the Global South) at a significant disadvantage in their advocacy (UNIDIR 2017: 22). It is therefore significant that the majority of working definitions under discussion have been put forth by Northern Developer states: in a 2017 primer report, 'The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches' UNIDIR highlights definitions from the Netherlands, France, Switzerland, the US and the UK but does not include the perspectives of any Global South states (UNIDIR 2017). While the relative clout of these actors compared to many smaller, less-developed Southern states likely influences their ability to garner support for such definitions, it is also necessary to understand that the comparative capacity of these states to create a technical definition of this sort is inhibited by their status as Observers.

Written by Alena Zafonte

This fact makes the participation of non governmental organizations and civil society all the more important as these third sector actors can provide platform to Observer state concerns. The definition put forth by the International Committee of the Red Cross (ICRC) has garnered significant discussion as it takes a functionalist approach rather than a technical one (UNIDIR 2017). The definition states that an autonomous weapons system is "any weapons system with autonomy in its critical functions. That is, a weapon system that can select [...] and attack [...] targets without human intervention." (ICRC 2016: 1) Similar definitions have been offered from NGOs around the world, including UK NGO Article 36 which is credited with bringing the concept of 'meaningful human control' into the discussion (UNIDIR 2017; Article 36 2017). Organizations such as the ICRC and other non-state actors maintain that "human control over attacks is inherent in, and required to ensure compliance with, the IHL [international humanitarian law] rules of distinction, proportionality and precautions in attack" thus giving voice to the concerns of off-ignored Observer states (UNIDIR 2017: 26; Private Meeting 5 2018).

Another important point is that of the definitions highlighted by UNIDIR, that put forward by France is the only one to specify the targeting of humans by these autonomous systems (UNIDIR 2017). In doing so, France implies that it envisions regulations on autonomous systems "not applying to anti-material weapons, countermeasure systems, or non-kinetic systems." (UNIDIR 2017: 25) This signals that, although France supports limitations within the CCW framework, it remains a Developer state rather than an Observer and thus the Observer-Developer framework has served its purpose (Garcia 2015).

Legal Accountability

As demonstrated throughout history, in times of conflict "accountability is essential in international law to deter and prevent violations and thus to protect potential victims of human rights abuses, war crimes, and the like." (Hammond 2015: 662) It is also true that the absence of such accountability would render just war principles entirely ineffective, thus fostering cultures of war that would undoubtedly challenge existing international laws and norms (Hammond 2015; UNIDIR 2017). Several scopes of accountability have been proposed, but perhaps the most relevant are individual accountability and state accountability.

Scholars have identified two principal individual actors that may, in theory, be held accountable for LAWS action. The first is the commanding officer: under the principle of 'command responsibility' this individual can be held liable for the crimes of someone in their command if there is a defined superior-subordinate relationship or in the event that they have been notified concretely or constructively of an impending crime and fail to take action prevent it (Hammond 2015; UNIDIR 2014). While there are instances in which commanding officer responsibility would be easily identifiable (for example, if an officer should knowingly authorize an attack which would violate international law), it is important to note that "it is not clear, however, that commanders will consistently know of these types of risks" given that it is likely that they were not involved in the designing or programming of any autonomous weapons under their instruction (Hammond 2015: 665; UNIDIR 2017). Holding a commanding officer "responsible for AWS action that [they] could neither control nor foresee would thus go beyond the traditional scope of command responsibility." (Hammond 2015: 665) Furthermore, delegating a task to an autonomous system where the system must dictate how an order will be carried out without human input or intervention provides "less direct causality." (UNIDIR 2017: 14) This may ultimately serve to "flatten out some of the hierarchical structures in military organizations," a fundamental principle of major military powers, making command officer responsibility a less than desirable option for Developer states. Such legal attribution would ultimately move more decision making power to "non-human agent[s] that cannot be held accountable," thus also weakening international law (UNIDIR 2017: 14).

The other individual who may theoretically be held accountable is the designer or manufacturer— this approach would treat crimes committed by autonomous systems as a legal accident, and "manufacturers would be required to pay for any damages caused to compensate victims or their families." (Hammond 2015: 665) Though this would incentivize manufacturers to make these products as safe and reliable as possible, designer/manufacturer accountability has several pitfalls: manufacturers are seldom held responsible for the ill effects of design flaws as they can notify clients that such issues may occur, and even if such legal action were taken against them, that responsibility (and the cost of associated legal fees) would fall solely on the individual which suffered the damages (Hammond 2015: UNIDIR 2017). It is simple enough to understand how in Developer state legal systems such as the

Written by Alena Zafonte

United States, this approach would fail the civilian victims yet benefit the state. Unfortunately, in Observer states there would likely be even greater barriers to having such cases heard, meaning this approach would have dire implications for human rights.

Similar issues arise when considering the possibility of state accountability: by some accounts, state accountability is "normatively and theoretically superior to individual accountability," but its practicality remains uncertain at the present time (Hammond 2015: 658). This uncertainty stems from a lack of unified legal conclusions regarding how such accountability would be attributed as there are two options for imposing state accountability: victim states (those whose citizens were affected) could take action against perpetrators in the International Court of Justice (ICJ), or individual victims could bring charges against the aggressor states (Hammond 2015; UNIDIR 2017). In considering the ICJ route it is clear that, though the ICJ has a broad subject matter jurisdiction, the strong limitations on personal jurisdiction would "likely obstruct its power to hear AWS [autonomous weapons systems] disputes." (Hammond 2015: 657) Furthermore, there is concerning precedent found in the 1986 ICJ ruling in the Case Concerning Military and Paramilitary Activities in and Against Nicaragua: Nicaragua brought charges against the United States in the ICJ related to the US' involvement in acts committed by the Contra, a US-backed rebel group which received American funding, training and supplies (Nicaragua v. United States of America 1986; Human Rights Watch 2016). The ICJ ruled that for state accountability to apply it would need to be "proved that State had effective control of the [...] operations in the course of which the alleged violations were committed."(Nicaragua v. United States of America 1986; Human Rights Watch 2016) In the absence of such evidence the United States was not held accountable (Nicaragua v. United States of America 1986: 65.) The ICJ also "lacks an enforcement mechanism" to adequately oversee remedies for potential cases and many Developer states that may eventually employ LAWS do not fall under the ICJ's compulsory jurisdiction (Hammond 2015: 657; Private Meeting 5 2018). While suits brought against states by individuals lack these pitfalls, the victims of such violence are "oft-impoverished and poorly situated" and therefore are without the means to pursue such justice (Hammond 2015: 657).

While there is no clear consensus on which approach to accountability best addresses the risks posed by LAWS, most Developer states agree that a degree of 'meaningful human control' (MHC) is necessary in the employment of such technologies, thus lending itself to the concept of individual accountability (UNIDIR 2014: 2; Article 36 2017). Though meaningful human control as developed by Article 36 explicitly referred to MHC over individual attacks, the phrase, and related conceptual iterations, have been adapted to the general operations of autonomous weapons systems (UNIDIR 2014; Human Rights Watch 2016; Article 36 2017). The United States and Israel, both Developer states, have "advocated for using the term 'appropriate human judgement' rather than meaningful human control" as MHC can be seen as redundant with general technological development practices (Human Rights Watch 2016:1). There is precedent for such language as the Mine Ban Treaty prohibits victim activated mines but not those that are command detonated— this distinction is made on the grounds that victim activated mines (those without human control) pose a greater risk to noncombatants (Human Rights Watch 2016).

Observer states generally agree that a certain degree of human control is required in the employment of these systems; the Non Aligned Movement has pushed publicly for the CCW to move swiftly to "develop new international law on fully autonomous weapons." (Campaign to Stop Killer Robots 2018: 3) The contributions to the technical aspects of this discussion have been notably weaker from the Observer states as there is lesser capacity for institutional knowledge about the opportunities for such human control within the development and deployment processes (Private Meeting 1 2018; Private Meeting 5 2018). However, the NAM position favoring human control aligns with the general wish of Observers for international legal limitations on autonomous weapons technologies.

Outlook and Conclusions

By examining definitional approaches and the discussion of legal attribution in the CCW through a Developer vs Observer framework, it has become clear that discrepancies in the technological development of LAWS have contributed to the Convention's stagnation. This structure has also provided insight into the role of the ever present North-South divide in these talks: the gap between advancement between the North and South has fostered a rift in the relative ability to contribute to legal discussions on these emerging technologies in multilateral forums. This unfortunate reality sheds light on a fundamental principle of this paper's inspiration: as demonstrated, most policies,

Written by Alena Zafonte

norms and research come from Developer states which are almost exclusively in the Global North, and yet the South is consistently more heavily impacted (Private Meeting 3 2018). This means that not only are the security concerns and capacities of Observer states different than Developer states', but their ability to meaningfully participate in the discussion is also impacted. The dissonance between Observer and Developer states therefore sits high amongst the many confounding factors precluding effective CCW action on LAWS.

In examining definitional and attribution approaches these distinctions only become more clear. As the NAM continues to pursue preemptive limitations on LAWS and Developer states continue to advance their technological pursuits over protections for human security, it is unfortunately likely that the already wide gap between these two global perspectives will widen. Furthermore, "developments will not necessarily be under the control of militaries," meaning that this discussion will, at some point, become further nuanced in order to account for nonstate actors such as multinational corporations or, perhaps more frighteningly, armed groups (UNIDIR 2017:8). Though civil society actors such as the Campaign to Stop Killer Robots and Article 36 remain active in championing a ban or other limitations, their efforts will likely remain unsuccessful until technological discrepancies shrink to the point at which the benefits of LAWS are relatively equal between Observer and Developer states: without equal benefit, the Observers will always favor regulation so as to protect their citizens from harm. Unfortunately, it is unlikely that any significant progress will be made in these talks in the near term as the discussion on definitions of autonomy and appropriate legal attribution remain somewhat convoluted.

This paper began by asserting that one's position on any particular policy depends on "what one is concerned about," and therefore concludes by emphasizing that the North-South division that has historically been present, while undeniably still at play, is not sufficient on its own to explain different positions on LAWS in the CCW (UNIDIR 2017: 1). As is clear, it is effectively accompanied by an additional classification such as the Observer vs Developer state statuses outlined in this paper as this allows the nuance necessary to understand the impact of the North-South relationship on these talks; further research in this area is therefore crucial.

Bibliography:

Article 36 2017, *Autonomous weapon systems: Evaluating the capacity for 'meaningful human control' in weapon review processes,* 13-17 November, viewed 12 June 2018 http://www.article36.org/wp-content/uploads/2013/06/Evaluating-human-control-1.pdf

Burri, T 2016, 'The Politics of Robot Autonomy', European Journal on Risk Regulation, Vol 2., pp. 341-360. https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/div-classtitlethe-politics-of-robot-autonomydiv/20FC3CF4F81602B4E7C97E68E4A0B516

Campaign to Stop Killer Robots 2018, *Report on Activities: November 2017*, 26 February 2018, viewed 15 June 2018, https://www.stopkillerrobots.org/wp-content/uploads/2018/02/CCW_Report_Nov2017_posted.pdf

Garcia, D 2015, 'Governing Lethal Autonomous Weapon Systems', Ethics and International Affairs, 13 December 2017, pp. 1-6 https://www.ethicsandinternationalaffairs.org/2017/governing-lethal-autonomous-weapon-systems/

Hammond, D 2015, 'Autonomous Weapons and the Problem of State Accountability', Chicago Journal of International Law, Vol. 12 No. 2, pp. 652-687.

Human Rights Watch 2016, *Killer Robots and the Concept of Meaningful Human Control*, 11 April, viewed 16 June 2018 https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control

Institute for Economics & Peace 2018, *Global Peace Index 2018: Measuring Peace in a Complex World*, June, viewed 15 June 2018 http://visionofhumanity.org/app/uploads/2018/06/Global-Peace-Index-2018-2.pdf

Marr, B 2018, 'Weaponizing Artificial Intelligence: The Scary Prospect of Al-Enabled Terrorism', 23 April, viewed 16 June 2018 https://www.forbes.com/sites/bernardmarr/2018/04/23/weaponizing-artificial-intelligence-the-scary-

Written by Alena Zafonte

prospect-of_ai-enabled-terrorism/#2eea3de177b6

Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14.

Private Meeting 1, 11 June 2018, Geneva, Switzerland

Private Meeting 2, 5 June 2018, Geneva, Switzerland

Private Meeting 3, 29 May 2018, Geneva, Switzerland

Private Meeting 4, 30 May 2018, Geneva, Switzerland

Private Meeting 5, 31 May 2018, Geneva, Switzerland

'Profile: Non-Aligned Movement' 2009, BBC, 7 August, viewed 14 June 2018 http://news.bbc.co.uk/2/hi/2798187.stm

United Nations Institute for Disarmament Research (UNIDIR) 2018, *The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence*, viewed 15 June 2018, http://www.unidir.org/publications

United Nations Institute for Disarmament Research (UNIDIR) 2014, The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, viewed 15 June 2018, http://www.unidir.org/publications

United Nations Institute for Disarmament Research (UNIDIR) 2017, *The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches*, viewed 14 June, 2018 http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf

Views of the International Committee of the Red Cross (ICRC) on autonomous weapons systems 2016, International Committee of the Red Cross (ICRC), 11 April, viewed 12 June 2018 https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system

Wolvers, A, Tappe, Salverda, Schwarz, 'Concepts of the Global South: Voices From Around The World', Global Studies Center, University of Cologne, Germany http://kups.ub.unikoeln.de/6399/1/voices012015_concepts_of_the_global_south.pdf

World Economic Forum 2018, *The Global Risks Report 2018*, viewed on 16 May 2018 http://www3.weforum.org/docs/WEF GRR18 Report.pdf

Written by: Alena Zafonte Written at: Northeastern University Written for: Professor Denise Garcia Date written: June 2018