The role of the private sector in the governance of autonomous weapon systems:

A principal-agent perspective

--- Draft for We Robot 2017 – Please do not cite without permission ---

Sophie-Charlotte Fischer

Swiss Federal Institute of Technology (ETH Zurich)

Center for Security Studies

Haldeneggsteig 4, CH-8092

sophie.fischer@sipo.gess.ethz.ch

Introduction

Arms Control was a crucial backbone of international security policy during the Cold War. More than half a century later, the landscape of security challenges has transformed but the need to tackle problems raised by new means and methods of warfare did not lose its significance. Currently, states are confronted with an increasing number of emerging technologies, scientific techniques and substances, which challenge the adaptability and conceptual foundations of established arms control regimes (Clapper, 2016; McCreight, 2016; World Economic Forum, 2015). These new developments are likely to affect strategic stability and jeopardize the protection provided by the principles and rules of international humanitarian law.

One of the greatest emerging challenges is the increasing autonomy in the critical functions of weapon systems, made possible by technological advances in areas including artificial intelligence, sensor technology and computing power. The possible development and deployment of fully autonomous weapon systems (FAWS), that could select and engage targets without any human intervention, raise difficult legal and ethical question (ICRC, 2015).

Although the issue has been taken up by states under the United Nations Convention on Certain Conventional Weapons (CCW) in Geneva, the prospects for an internationally agreed protocol to ban or restrict FAWS in this framework are uncertain (Boulanin, 2016a). Moreover, the deliberations in Geneva but also the academic literature on robot arms control have so far largely neglected that key technological components of FAWS are dual-use (World Economic Forum, 2017; Biontino, 2016; Francois, 2015), which generally decreases the prospects for and effectiveness of hard law measures.

In order to contribute to a comprehensive discussion, this paper widens the analytical scope by adopting a de-centered perspective of arms control. Through a global security governance framework, the role of private technology companies in the governance of autonomous weapon systems is scrutinized. I argue that the lead of private technology companies in innovation and development of dual-use technologies in robotics and artificial intelligence (AI) merits an analytical focus on these actors. Through the principal-agent problem, a model originally developed in economics, I

demonstrate how the lead in technology innovation and development endows private technology companies also with an augmented regulatory capacity vis-à-vis governments. The paper concludes with a discussion of the benefits and shortcomings of existing private measures in the Information and Communications Technology (ICT) industry regarding the governance of autonomous weapon systems. I conclude that private governance measures are no silver bullet for the governance of autonomous weapon systems but can be very beneficial, if they are embedded in a public oversight-framework.

The paper is sub-divided into three main parts. In the first part, I provide an overview of the current state of robot arms control, demonstrate the limits of established arms control regimes and argue why there is a need to discuss the issue in a global security governance framework. In the second part, I investigate the role of private technology companies in the governance of autonomous weapon systems. In the two subsections of this part, I examine the two-fold function of private technology companies as driving force behind the development of increasingly autonomous systems and as regulator of this technological evolution. In the third part, I discuss the implications of existing private governance measures for the governance of autonomous weapon systems. I also provide suggestions for how public actors can harness the benefits of these private governance measures while counteracting some of their shortcomings.

The main contribution of this paper is threefold. First, I challenge the prevailing view, that arms control is an inherently state-centric practice. By bridging the literatures on robot arms control and global security governance, I shed light on the contributions of other actors to the governance of autonomous weapon systems. Second, I unfold and highlight the crucial role of private technology companies in the development and governance of autonomous weapon systems, which has been neglected in the academic discourse and at the CCW in Geneva so far. Third, I provide recommendations for how public actors can shape emerging private governance measures in the ICT industry.

Autonomous Weapon Systems and Challenges to Existing Arms Control Regimes

The world is currently on the cusp of a third revolution in military affairs, characterized by the transfer of a growing variety of military tasks to machines. The areas of application for robotic systems currently cover mainly "dull, dirty and dangerous" jobs. In the future, however, these could possibly also include the application of lethal force without human interference (Haas, 2014). Although fully autonomous weapon systems, that could select and engage targets without human intervention, do not yet exist, leading researchers in robotics and artificial intelligence argue, that this development is practically feasible within years (Future of Life Institute, 2015).

The primary foundation for machine autonomy is artificial intelligence (AI) (Defense Science Board, 2016, p.5). Simplified, AI can be defined as "the capability of machines to perform complex tasks such as decision-making and perception that were normally restricted to humans" (Cummings, 2017, p.2). In fully autonomous weapon systems, AI would enable the machine to reason probabilistically based on a set of inputs, compose different courses of actions and select and execute the best option without human interference at any of these stages (Cummings, 2017; Defense Science Board, 2016). Currently, FAWS are a 'virtual reality'[1] as they do not yet exist. However, scientific advances in AI enable increasing levels of autonomous functionality (Defense Science Board, 2016). Functions of weapon systems to which autonomy can be "attached" include a) mobility b) health management c) interoperability d) battlefield intelligence and e) the use of force (Boulanin, 2016b, p. 7&8).

The ongoing scientific developments in AI have sparked a debate on whether artificially intelligent machines should be allowed to execute lethal force without human intervention. While many critics question, that FAWS will ever be able to act in conformity with international humanitarian law, others argue that even if this was technologically feasible, machine killings would still be incompatible with the protection of human dignity (ICRC, 2015). Moreover, the perceived benefits of FAWS, such as the increasing speed of decision-making processes and the augmented survivability of

[1] The term has been coined by the German sociologist Ulrich Beck, but has been also used in relation to FAWS in a recent study by the Swedish Think Tank SIPRI (Boulanin, 2016a, p.3).

the systems (Lewis, 2016) could catalyze a new arms race between states (Wallach & Allen). Furthermore, if critics are to be believed, autonomous weapon systems would lower the threshold to initiate wars by replacing soldiers on the battlefield (Gubrud, 2016).

Arms Control or Security Governance?

Following an influential campaign by a coalition of civil society organizations[2], states took up the issue of FAWS in 2013 under the United Nations Convention on Certain Conventional Weapons. The CCW provides a framework for state parties to restrict or ban new types of weapons that are considered "to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately" in the form of an annexed protocol to the Convention (UNOG, 2017). In the years 2014 to 2016, three informal meetings of experts were held at the UN in Geneva with the aim to discuss emerging technologies in the area of FAWS and to develop a shared understanding of them. This groundwork cumulated in the establishment of a Group of Governmental Experts (GGE) at the CCW Review Conference in December 2016 and thereby initiated the formal negotiation process (UNOG, 2017b).

The past meetings of experts, and the recently established GGE are undoubtedly valuable platforms to find common definitions of core concepts in the debate, to clarify national positions and to build confidence among states. However, after three years, the process, is still at an early stage and the outcome remains uncertain (Boulanin, 2016a). So far, the parties to the CCW could not agree on a common definition of what constitutes an autonomous weapon system and are therefore still in the process of understanding the subject under discussion. Moreover, competing narratives on what regulatory approach would be most suitable have already emerged although the official negotiations have yet to start (Dunn-Cavelty, Fischer, Balzacq, 2016).

In this contribution, I argue, that the narrow focus on the regulatory capacity of states and the development of traditional arms control measures in the framework of the CCW is insufficient to effectively deal with the intricate challenges posed by autonomous weapon systems. This argument is based on mainly three observations

[2] For further information on the work of civil society organizations on FAWS and their history see www.icrac.net and https://www.stopkillerrobots.org/

that highlight the unique technological features of autonomous weapon systems and as well as the current limits of hard law measures.

First, a critical feature of autonomous weapon systems that has been largely overlooked at the CCW as well as in the scholarship on robot arms control, is that major technological components of FAWS are dual-use and can thus have both military and civilian applications (Munich Security Report 2017; Biontino, 2016; DoD, 2016).[3] The enormous potential of autonomous systems for the civil and military sphere, makes them particularly resistant to preventive hard law measures, as they could limit their development and use in both spheres. Several state parties to the CCW, especially high-tech exporting countries like Japan and South Korea, have repeatedly voiced their concerns that a protocol on FAWS could affect the civilian innovation and development process of autonomous systems (Youngjip, 2015; Nagayoshi, 2015). At the current stage of development, where the full benefits but also potential risks are not entirely foreseeable yet, it is challenging to foreground ethical, legal and security considerations in state's military and economic planning (Wittes & Blum, 2015; Collingridge, 1981).

Second, the creation of new multilateral regimes in international fora is stagnating and increasingly time- and resource- intensive (Pauwelyn, Wessel & Wouters, 2012). Within the CCW, negotiations are prolonged and complicated through the consensus principle, which makes the establishment of a multilateral regime in this context unlikely in the foreseeable future. Moreover, a lengthy regulatory process is quickly outpaced by the further accelerating technology development (Marchant, 2011; Moses, 2007). This dynamic inevitably compromises the desired effects of preemptive legal instruments on autonomous weapon systems, that are discussed in Geneva.

Third, the dual-use character of autonomous systems components makes verification and therefore also industry cooperation a crucial feature of an effective arms control regime. The CCW, however, lacks enforcement and verification mechanisms (Arms Control Association, 2007), which would considerably weaken the effectiveness of a future protocol on FAWS from its onset.

The current legal void and limitations of hard law measures against the

[3] An exception to this is a side-event on private sector perspectives on FAWS, sponsored by the Permanent Representation of Germany to the Conference on Disarmament during the CCW meeting of experts in 2016 (see Biontino, 2016).

progressing technology development, create the need to consider and scrutinize the potential of alternative approaches. I therefore argue that the search for measures to govern autonomous weapon systems needs to be expanded to the regulatory capacity of non-state actors. To overcome the limitations of the existing scholarship on robot arms control, which has primarily focused on the problem-solving capacity of states, I adopt a de-centered perspective of arms control through a global security governance framework.

A security governance framework has the aim to extend the analytical scope to different actors and instruments that influence the development and control of autonomous weapon systems beyond the state. The value of a global security governance approach lies in its ability to accommodate some of the fundamental changes to the ways in which international security has been coordinated, provided and regulated after the Cold War (Biersteker, 2016). Moreover, it allows "to observe and analyze phenomena and trends that might go unnoticed from traditional approaches to security policy" (Ehrhart, Hegemann & Kahl, 2013, p.147).

Global Security Governance as Analytical Framework

The origins of global security governance can be traced back to a larger trend that has profoundly altered the ways of governing after the end of the Cold War. Traditionally, the authority to take collectively binding decisions on both, the national and international level was restricted to public actors, referring either to national governments or their interactions through intergovernmental institutions (Wolf, 2008, p.225). However, forces of globalization, such as privatization and trade liberalization, which accelerated during the 1990s, fundamentally reshaped the landscape of challenges that states were confronted with (Krahmann, 2003). The rising number and complexity of problems from diverse issue areas that were not confined to national borders, revealed the limits of states' problem-solving capacity and resources (Fukuyama, 2016). Yet, the resulting policy void also brought attention to the unique resources that actors other than the state could contribute to public policy making. Consequently, authority for certain tasks that were formerly administered exclusively by states, was assumed fully or in part by international and supranational organizations, regional and local institutions and most notably non-state and private actors (Daase & Engert, 2009).

The term governance is therefore often used in opposition to the more traditional concept of government and describes at its core the emergence of fragmented, "multilateral and non-territorial modes of regulation, with the participation of private and non-governmental actors" (Parker & Braithwaite, 2003, p.506; Acharya, 2016). At times, the activities of different governance arrangements that target the same issue area can coexist, overlap or possibly even contradict each other (Biersteker, 2016, p. 428). However, as Stoker (1998, p.17) stresses, the aim of governance remains "to create the conditions for ordered rule and collective action". What has therefore changed through the fragmentation of authority and new governance arrangements is not the aspired outcome but rather the process of attaining it.

Corporations are one group of actors, which gradually assumed the position of public bodies in rule-making across a variety of industries (Haufler, 2001, Ronit, 2011).[4] According to Biersteker (2016, p. 431), corporations can obtain authority in global governance through three different mechanisms "when states delegate it, enable it or passively allow it to develop". Private governance measures generally take the form of non-binding, voluntary soft-law measures and range from codes of conduct and standard setting schemes to coordinated lobbying efforts and recommendations by private ethics boards (Cutler et al., 1999; Knowles, 2012, p.59). The objective of these voluntary measures is to "influence or control behavior for the benefits of the organization itself and for the communities in which they operate" (Gordon & Miyake, 2000, p.3). While some companies unilaterally develop private governance measures, others band together with market competitors as industry consortia and make commitments collectively. These new regulatory activities were theorized by some governance scholars as the emergence of "private authority" in international affairs (Hall & Biersteker, 2002; Cutler, Haufler & Porter, 1999).

The emergence of new modes of governance including private actors has also affected the area of global security. However, the fragmentation of authority in this domain is particularly noteworthy as the provision of security, an essential public good, has always been considered as a core prerogative of states and even as justification for their monopoly on violence (De Neves, 2010, p.219). Therefore, as Dunn-Cavelty

[4] Diverse industries, such as health care, textile, telecommunications and mining have used self-regulatory processes to govern industry practices. See for examples Cutler, Haufler & Porter (1999).

and Suter (2009, p.181) note, “it is an extremely delicate matter for governments to pass on responsibility in this area to the private sector”. This, however, makes it even more important to understand how private actors shape security policy and whether private governance measures can ultimately reduce or further augment global security risks (Daase & Engert, 2009).

In the academic and public discourse on autonomous weapon systems, it has gone largely unnoticed that there are currently several overlapping governance arrangements and soft-law instruments that directly and indirectly shape the further development and use of the technology. These include *inter alia* the Global Initiative for Ethical Considerations in the Design of Autonomous Systems established by the Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA, 2016). Moreover, an epistemic community of AI researchers recently developed the “Asilomar AI Principles”, 23 guidelines targeted at AI researchers and lawmakers to ensure the ethical and beneficial use of AI (Future of Life Institute, 2017). However, another important group of actors, which has received surprisingly little attention in the existing scholarship on robot arms control but also during the discussions on FAWS at the CCW is private technology companies.

The key role of private technology companies in the innovation and development of autonomous systems, its regulatory capacity vis-à-vis national governments and intergovernmental organizations as well as the potential benefits and limitations of existing private governance measures, merit the focus on these actors.

The Role of the Private Sector in the Governance of Autonomous Weapon Systems

Arms control, as one of the mainstays of international security during the Cold War, has always been administered through treaty-based intergovernmentalism, covering bilateral and multilateral measures. Traditionally, the access of corporations to intergovernmental bodies like the UN has been limited, as they were perceived as actors whose essential goals are primarily profit-driven and impossible to align with public interests (Ronit, 2011, p. 79). Moreover, as pointed out by Carr (2016, p.43) the private sector is considered as being generally averse to accepting liability or responsibility for security issues.

In established areas of arms control, transnational corporations have developed

an image as "originators of regulatory problems rather than problem solvers" (Wolf, 2008, p.223). This negative view originates mainly from the ties of several businesses from countries such as Germany, South Africa and Singapore with the AQ Khan proliferation network, which sold dual-use goods and technologies for the build-up of nuclear programs to several authoritarian regimes worldwide (Broad, Sanger & Bonner, 2004). The lacking support of the biotech industry for the Biological Weapons Convention (BWC) further tarnished their reputation and complicated cross-sector cooperation on arms control (Busch & Pilat, 2017, p.29). The "Wiesbaden Conference", a yearly industry forum initiated in 2012 and facilitated by the UN Office for Disarmament Affairs, is an attempt to better sensitize businesses for their role and obligations in averting the proliferation of dual-use technologies under UN Security Council Resolution 1540 (UN, 2012).[5] Summing up, corporations have hitherto been challenging rule-takers rather than supportive rule-makers in arms control.

This role has changed with the emerging issue of autonomous weapon systems. Here, private technology companies have a twofold function; first, as the driving force behind the innovation and development of dual-use technologies pertinent to FAWS and second, as a regulator of this technological revolution.

The Reversed Innovation Process of Dual-Use Technologies

Traditionally, dual-use items were developed from defense funds exclusively for military application and transferred only at a later stage to the civil realm, popular examples being GPS systems or transistors. Yet, after the Cold War, rapid growth of commercial markets on the one hand and declining defense budgets as well as faltering procurement procedures on the other hand, deprived government sponsored defense projects of the lead in research and development (R&D) (Bitzinger, 2009). Facing the two-fold challenge "to win the new global economy and safeguard our national security", the Clinton-Gore administration implemented the so-called Technology Reinvestment Project (TRP) in 1993 (Clinton, 1993, p.2110; Stowsky, 1996). One aim of the initiative was to foster the development of commercial technologies with potential military spin-offs through complementary government

[5] UN Security Council Resolution 1540 was adopted unanimously in 2004 and establishes the obligations to develop and enforce appropriate legal and regulatory measures against the proliferation of weapons of mass destruction under Chapter VII of the UN Charter. The full text of the resolution can be viewed here http://www.un.org/en/ga/search/view_doc.asp?symbol=S/RES/1540(2004)

investment in R&D (Stiglitz & Wallstein, 1999, p.52). The administration believed that "only the private sector has the abilities to manage the complex processes of developing new technologies and bringing them to the market" (Clinton in Carr, 2016, p.47) while government continues to play a vital role in supporting these processes.

Today, the consequences of this shift have come into full effect. Dual-use technologies, that are relevant to the development of autonomous weapon systems are predominantly driven by commercial needs and led by the private sector.[6] Private technology companies have significantly more capital at their disposal, enabling dynamic innovation processes and a faster introduction of new technologies on the market (Francois, 2015, p.12; Ronit, 2011, p.75). The lead of private sector innovation in the relevant industries is expected to further extend as the most qualified engineers from universities are drawn to the commercial rather than public sphere (Cummings, 2017, p.8; Hernandez & King, 2016).

Dual-use technologies that are central for the development of autonomous weapon systems are different applications of AI, as they enable increasing levels of autonomy in key functions like decision-making and perception. Hardware components, such as sensor technology, play an important role in supporting these functions and executing commands. The key innovations in AI come from the ICT and automotive industry, which develop the technologies for commercial purposes (Boulanin, 2016c, p.26). Regarding the ICT industry, for example, the Munich Security Report 2017 (p.56) notes that "*Google, Amazon, Microsoft* and *Intel* alone spend more than USD 50 billion a year on digital innovation with dual-use". Commercial technologies with dual-use can be adopted for military purposes through three different avenues. While many of them can be adopted off-the-shelf, others can be implemented after slight modifications. The third option is that the R&D efforts underlying commercial technologies serve to improve the military's internal development of autonomous capabilities (Boulanin, 2016c, p.26).

The implications of this shift in innovation and development capacity to private technology companies are manifold. While their R&D activities allow militaries to innovate despite of shrinking defense budgets, it also increases their dependency on

[6] This pattern is best developed in Western high-tech states. It also applies to other emerging dual-use technologies including advancements in biotechnology, nanotechnology and cybertechnologies, see for example World Economic Forum, (2015).

the private sector and transforms military acquisition processes (Defense Science Board, 2016; Lynn, 2014). Moreover, dual-use technologies have become more "democratic" as access to them is not restricted to commissioning governments any longer (Wittes & Blum, 2015). This extended accessibility, allows states to catch-up faster with the military capabilities of competitors. In addition, it also augments proliferation concerns as non-state actors with a malicious intent could acquire increasingly autonomous systems more easily (Horowitz, 2016).

Another crucial implication that has been neglected in the academic literature so far, is that the shift in innovation and development has increased the responsibility of the private sector for potential risks from emerging technologies. Moreover, it has augmented the regulatory-capacity of private sector companies in the governance of emerging dual-use technologies.

The Principal-Agent Problem

The previously outlined shift, together with the legal structure of corporations, have altered the regulatory capacity of private technology companies vis-à-vis governments. The principle-agent problem, a model that evolved from agency-theory in economics, offers important insights into the resulting dynamic. It can serve to "map" the relationship between governments and private technology companies and shed light on why private actors can and do introduce private governance measures, usually before public measures are introduced. Before applying the model to the case of the governance of autonomous weapon systems and discussing the resulting implications, the basic tenets of principal-agent models are briefly outlined.

Principal-Agent (P-A) models were originally developed in economics in order to explain how and why contracted agents engage in behavior that is undesired by contracting principals (Grossman & Hart, 1983). For example, a broker may not always act in the interest of a customer. P-A models theorize the distinct interests of principals and agents as well as the incentives and likelihood of an agent's autonomous action. An asymmetry in information is viewed as a major problem that enables autonomous action by the agent and at the same time limits the principal's control (Pollack, 1997). In International Relations (IR) P-A models emerged as a response to Neo-Realism (Waltz 1979; Mearsheimer 1994) and Neo-Liberalism (Krasner 1983; Keohane 1984) to theorize and explain autonomous action by international organizations (Nielsen &

Tierney, 2003; Pollack, 1997). Meanwhile, it has been applied to a range of other IR issues such as the regulation of private security companies in contemporary conflicts (Cockayne, 2007).

In case of the governance of autonomous weapon systems, I use the principal-agent problem to "map" the relationship between the state and private sector, designating the state as principal (P) and private technology companies as agents (A). In this constellation, and different to the classic principal-agent problem, authority has not been formally transferred by (P) (who has the democratic legitimacy to regulate) to (A) but delegation was rather implicitly and passively delegated through first, the current regulatory vacuum (Biersteker, 2016, p. 431) and second, the shift in innovation and development of relevant dual-use technologies.

Moreover, the shift in the technology innovation and development process, as outlined before, has created a substantial asymmetry in information between (P) and (A). Although (P) has shifted the task to develop dual-use technologies to (A) after the end of the Cold War, (A) has thereby established a quasi-monopoly in information on the development of dual-use technologies vis-à-vis (P) (Aven & Renn, 2010, p.49). This information is crucial to approximate the further development trajectory of the technology and to make informed decisions on this basis about whether and how it should be regulated.

Furthermore, due to its legal structure as corporation and its accountability that is limited to shareholders, (A) is much more flexible in its actions. (A) can thus introduce private governance measures at a much faster pace than (P) can implement binding regulations. This dynamic is augmented when states must act through intergovernmental bodies, which requires decision-making by unanimity like the CCW (Pitofsky, 1998; Marchant, 2011).

Lastly, (A) and (P)'s interest profiles and incentives for regulation differ. (A) is primarily concerned with developing and selling competitive and innovative technologies. On the other hand, (A) is also interested in keeping a good reputation among its client base (Haufler, 2001) and possibly, to act socially responsible (Matten & Crane, 2005). (P) is also interested in the revenues from commercial technologies, the prestige from innovation (Collingridge, 1981) and in certain cases the strategic benefits from military applications. (P) therefore has different disincentives to regulate the development process and trade preventively as it would harm its own interests. On

the other hand, (P) is also accountable towards society to safeguard national security and possibly regulate risks from emerging technologies.

Existing Industry Initiatives

High-level research and management staff of leading companies in AI development, including inter alia *Microsoft, Apple, Tesla, Facebook* and *Google*, has spoken out against the development of offensive autonomous weapon systems in an open letter published by the Future of Life Institute (2015). Until today, there are no transnational regulations in place that govern the further development and potential uses of autonomous weapon systems. And although the CCW process was formalized through the establishment of a GGE, the prospects for a protocol on FAWS remain uncertain.

In the meantime, however, corporations from the ICT industry have set up a variety of private governance measures to support a safe and ethical development of AI. Most notably, five major Silicon Valley firms from the ICT sector, including *IBM, Microsoft, Amazon, Facebook* and *Google* have established the industry-consortium "Partnership on AI" in September 2016. The consortium aims at formulating best practices on AI technologies (Markoff, 2016). It also seeks to "advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society" (Partnership on AI, 2016). Recently, the Partnership on AI was joined by *Apple* as a sixth member. At the same time, the partnership announced the appointment of five additional members to its board of trustees from academia and civil society organizations (Hsu, 2017).

Other noteworthy unilateral initiatives include the internal ethics board of *Google's* AI powerhouse *Deepmind*, which was established as a condition by *Deepmind* when *Google* acquired the company in 2014. However, apart from the announcement of its establishment, no further information has been made publicly available on the composition or activities of the ethics board (Hern, 2017). In contrast, the Austin-based company *Lucid AI* has also established an ethics advisory panel, yet treats it considerably more open by providing information about its members and regular updates on its activities (Firth-Butterfield, 2016).

It is important to underline that even if the control of autonomous weapon systems is not the principal rational for why these initiatives were set-up, decisions taken by them will considerably influence their development and potential uses. The

companies will not only select their future customers and collaborators but also profoundly shape what will be technologically possible and ethically accepted.

Industry Self-Governance: Silver Bullet for the Governance of Autonomous Weapon Systems?

Although the portrayed initiatives are still in their initial phases, it is important to discuss already at this stage their potential benefits and limitations in the context of the governance of autonomous weapon systems. This is crucial given their influence on the further development and use of AI technologies in warfare and beyond. Drawing on the outlined cases of private governance measures, as well as insights from the business and management studies literature on global business regulation in other sectors, I argue that private governance measures in the ICT industry can be beneficial, yet only if they are embedded in an additional oversight framework provided by public bodies.

The need for an oversight framework

An oversight framework for private governance measures would have four main functions; first, monitoring continuously the commitments and corresponding conduct of private rule-makers, second, encouraging the industry-wide implementation of private-governance measures, third, facilitating the harmonization of measures across relevant industries and fourth, enabling a regular dialogue between private and public rule-makers to prevent the emergence of competing regulatory narratives.

An often-raised criticism of private governance measures is their missing democratic legitimacy (Scherer, Baumann-Pauly & Schneider, 2012; Vogel, 2008; Driver & Thompson, 2002). In the light of democratic standards, the legitimacy of private actors in rule-making is indeed problematic. Decision-makers in corporations, have not been elected by a constituency and it is unclear on whose behalf they act when making decisions that affect issues beyond the direct scope of the company. Accountability, in the case of corporate leaders, is limited to shareholders (Papadopoulos, 2013). This lack of democratic legitimacy is especially problematic in the case of AI. Due to their key role in innovation and development, decisions taken by influential companies such as members of the Partnership on AI, will considerably affect the further development trajectory of AI, which in turn will deeply influence

society beyond possible applications in warfare (Stilgoe & Maynard, 2017; Lincoln Kitman, 2016).

While these criticisms are certainly well founded and important from a perspective of democratic theory, they also ignore the realities of global governance and the current limits of states' rule-making capacity. While legitimacy can certainly not be transferred to corporate decision-makers in the same way as to elected politicians, it must be accounted for in different ways. The transfer of authority to private actors may be derived from a "credible commitment to basic norms or to general welfare or on the recognition of expertise and problem-solving resources" (Conzelmann & Wolf, 2007, p.100). The fulfillment of these preconditions alone, however, does not determine the quality of legitimacy. Legitimacy will ultimately depend on whether private technology companies follow generally accepted norms, and especially whether these commitments are reflected in their past, present and future actions. The conduct of private regulators therefore requires continuous and close monitoring by public actors to cultivate an accountability regime that reaches beyond the shareholders of the company.

A critical element of the legitimacy of private authority, is for example transparency in decision-making processes and regarding their outcomes (Vogel, 2008; Wolf, 2008). The publicly available information on the Partnership on AI, as well as the inclusion of civil society members with equal voting rights on the board of trustees is so far a positive indicator for transparency. The same applies to the open conduct of the company Lucid AI on the composition and activities of its ethics board. On the other hand, the secrecy around the ethics board of Deepmind, makes the initiative highly intransparent and thereby self-defeats its purpose. This lack of transparency contradicts the augmented responsibility of companies derived from their leading role in the development of AI. Public bodies and committed industry players therefore need to encourage a culture of responsibility by promoting best practices for transparency across the industry.

Another problematic trend in industry self-governance, that requires close attention and counteraction through an oversight framework, is the fragmentation of private governance measures. The implementation of different standards across an industry sector can at worst result in a "race to the bottom" where market forces favor those companies, which apply the lowest standards (Clark, 1995). An example of this

dynamic is the synthetic biology industry, where different standards for end-user controls have emerged, leading to the creation of two competing industry consortia. While one is promoting strict standards, the other one only committed to the application of weaker standards. Some other industry players decided to remain outside either consortium (Lok, 2009). In the case of the ICT industry, a public oversight framework could prevent fragmentation by facilitating an early harmonization of emerging measures before they have hardened and change has become difficult and comes at a high cost.

Given the possibility, that the international community will still develop an international treaty on FAWS, it is crucial to stimulate a regular exchange between public and private sector representatives. A great potential of the private sector lies in its role as the driving force of the technological evolution[7]. Information on the further evolution could greatly aid the development of a sustainable and effective international regime and prevent the emergence of competing private and public regulatory narratives. Therefore, states should include industry representatives in the negotiations in Geneva, while also using opportunities for exchange through the "Partnership on AI".

Benefits of Private Governance Measures

Apart from the discussed limitations, private governance measures also have benefits that international agreements often lack and that should therefore be harnessed. For example, private measures applied by globally operating companies can have a wider reach than certain intergovernmental treaties (Conzelmann & Wolf, 2007). The latter often suffer from lacking support of key countries that would be essential to promote the objectives of the treaty.[8] If a company commits to private governance measures, they apply to a companies' activities worldwide, independent of the geographical location of the companies' branches. Thereby, a culture of responsibility in the further development of AI could also be spread to key countries, which currently object transnational measures.

[7] See also the former German Representative to the Geneva Disarmament Conference Michael Biontino on this aspect http://www.genf.diplo.de/contentblob/4781866/Daten/6452979/20160412sideeventwefredeen.pdf

[8] The Mine Ban Treaty, also known as Ottawa treaty suffers for example from lacking support of key states like the United States, Russia and China. For further information see http://www.icbl.org/en-gb/the-treaty/treaty-status.aspx

Leading companies in the field, such as the members of the Partnership on AI have the unique ability to not only raise awareness for a critical issue but also to persuade other important industry players to join forces. The voluntary nature of private governance measures lowers the bar for other competitors to join and collectively develop and diffuse standards across the industry. Private rules may also be considered as more legitimate by the industry than public standards because they are technically appropriate and less intrusive (Papadopolous, 2013).

Lastly, companies can more flexibly adapt private governance measures, while the technology is still developing. As the evolution of certain technological features that require caution are not be foreseeable yet, it might be necessary to adapt current measures along the trajectory (Collingridge, 1981). Hard law instruments, in contrast, run the risk of being outdated quickly, when the technology develops differently than expected by the instrument's designers (Marchant, 2011). Re-negotiating or adapting an international treaty would be again notoriously difficult and time-intensive.

Conclusion

In this paper, I analyzed the role of private technology companies in the governance of autonomous weapon systems. By highlighting the dual-use character of technologies that enable increasing degrees of autonomy in weapon systems, I demonstrated the limitations of established arms control regimes and the regulatory capacity of states. These limitations and the progressing technology development, create a pressing need to consider the potential of other actors to contribute to the regulation of autonomous weapon systems. Through a security governance framework, I widened the analytical scope to include the regulatory capacity of non-state actors into the inquiry. At the center of this investigation were private technology companies due to their leading role in the innovation and development of key dual-use technologies for FAWS. I used the principal-agent problem, to demonstrate how the private sector's lead in innovation and development also augmented its regulatory capacity vis-à-vis state actors. In the final part of this contribution, I discussed the benefits and limitations of emerging private governance measures in the ICT industry. I concluded that an oversight framework provided by public bodies is necessary to exploit the benefits of private measures for the governance of autonomous weapon systems.

Due to the little available information on the recently established governance arrangements in the ICT industry, their analysis in this paper remained superficial. Future research will therefore have to reevaluate their legitimacy and contribution to the governance of autonomous weapon systems. Another limitation of this paper and an avenue for further research is a discussion of more concrete ideas for how the evolution and use of digital dual- use technologies like AI could be practically restricted. Innovation processes in the vast ecosystem comprised of companies and start-ups that develop AI, will be difficult to monitor and regulate in practice. Moreover, the proliferation of existing digital technologies is considerably complicated by the intangible nature of code.

At the time of writing, weapon systems with increasing degrees of autonomy are already an integral part of the military arsenals of several powerful countries (Haas, 2014). At the same time, robots are also becoming noticeably more pervasive in a variety of areas in the civil sphere. Pressure from both spheres is likely to increase rather than decrease and will continue to fuel the development of AI and other dual-use technologies in the field of robotics. Therefore, the search for innovative models to govern increasingly autonomous weapon systems for the benefit of society must be an even higher priority for scholars and policy makers in the future.

References

Acharya, A. (2016). The Future of Global Governance: Fragmentation May Be Inevitable and Creative. Global Governance, vol. 22, 453-460.

Arms Control Association (2007). *Convention on Certain Conventional Weapons (CCW) At a Glance*. Retrieved from https://www.armscontrol.org/factsheets/CCW

Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94(886), 687-709.

Aven, T. & Renn, O. (2010). *Risk Management and Governance: Concepts, Guidelines and Applications*. Berlin: Springer.

Bierstecker, T. (2016). Global security governance. In Dunn-Cavelty, M. & Balzacq, T. (Eds.) *Routledge Handbook of Security Studies (2nd Ed.)* (425-435). New York: Routledge.

Biontino, M. (2016). *Introductory Statement by H.E. Ambassador Michael Biontino, Permanent Representative of Germany to the Conference on Disarmament Geneva, 12 April 2016*. Private Sector Perspectives on the Development of Lethal Autonomous Weapons Systems. Retrieved from http://www.genf.diplo.de/contentblob/4781866/Daten/6452979/20160412sideeventwefredeen.pdf

Bitzinger, R. A. (2009). *The Modern Defense Industry*. Santa Barbara: ABC-Clio, LLC.

Boulanin, V. (2016a). Mapping the Debate on LAWS at the CCW: Taking Stock and Moving Forward. *Non-Proliferation Papers, No. 49, March 2016*. Stockholm: SIPRI.

Boulanin, V. (2016b). *Mapping the Development of Autonomy in Weapon Systems: A primer on Autonomy*. SIPRI Working Paper, December 2016. Stockholm: SIPRI.

Boulanin, V. (2016c). *Mapping the Innovation Ecosystem Driving the Advance of Autonomy in Weapon Systems*. SIPRI Working Paper, December 2016. Stockholm: SIPRI.

Broad, W., Sanger, D. & Bonner, R. (2004). A Tale of Nuclear Proliferation: How Pakistani Built His Network. *The New York Times (online)*. Retrieved from http://www.nytimes.com/2004/02/12/world/a-tale-of-nuclear-proliferation-how-pakistani-built-his-network.html?_r=0

Busch, N. & Pilat, J. F. (2017). *The Politics of Weapons Inspections: Assessing WMD Monitoring and Verification Regimes*. Stanford: Stanford University Press.

Carr, M. (2016). Public-Private Partnerships in national cyber-security strategies. *International Affairs*, 92 (1), 43-62.

Clapper, J. R. (2016). *Worldwide Threat Assessment of the US Intelligence Community*. Retrieved from https://www.dni.gov/files/documents/SASC_Unclassified_2016_ATA_SFR_FINAL.pdf

Clark, G. (1995). Global Competition and Environmental Regulation: Is the 'Race to the Bottom' Inevitable? In Eckersley, R. (Ed.), *Markets, the State and the Environment* (229-257). Berlin: Springer Link.

Clinton, W. J. (1993). Statement on the Technology Reinvestment Project. In *Public Papers of the Presidents of the United States: William J. Clinton.* Washington D.C.: Office of the Federal Register.

Cockayne, J. (2007). Make or buy? Principal-agent theory and the regulation of private military companies. In Chesterman, L. & Lehnhardt, C. *From Mercenaries to Market: The Rise and Regulation of Private Military Companies.* Oxford: Oxford Scholarship Online.

Collingridge, D. (1981). *The Social Control of Technology*. London: Palgrave Macmillan.

Conzelmann, T. & Wolf, K. D. (2007). The potential and limits of governance by private codes of conduct. In Graz, J. C. & Nölke, A. (Eds.), *Transnational Private Governance and its Limits*, (98-114). London: Routledge.

Cummings, M. (2017). Artificial Intelligence and the Future of Warfare. *Chatham House Research Paper*. Retrieved from https://www.chathamhouse.org/sites/files/chathamhouse/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings.pdf

Cutler, C., Haufler, V. & Porter, T. (Eds.). (1999). *Private Authority in International Affairs*. Albany: State University of New York Press.

Daase, C. & Engert, S. (2009). Global Security Governance: Kritische Anmerkungen zur Effektivität und Legitimität neuer Formen der Sicherheitspolitik. In Schuppert, G. & Zürn, M. (Eds.), *Governance in einer sich wandelnden Welt* (473-498). Wiesbaden: VS-Verlag für Sozialwissenschaften.

De Nevers, R. (2010). The Effectiveness of Self-Regulation by the Private Military and Security Industry. *International Public Policy*, 30(2), 219-240.

Defense Science Board (2016). *Summer Study on Autonomy*. Washington D.C.: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. Retrieved from https://www.hsdl.org/?view&did=794641

Driver, C. & Thompson, G. (2002). Corporate governance and democracy: The stakeholder debate revisited. *Journal of Management and Governance*, 6, 111-130.

Dunn Cavelty, M., Fischer, SC. & Balzacq, T. (2016). 'Killer Robots' and Preventive Arms Control. In Dunn Cavelty, M. & Balzacq, T. (Eds.), *Routledge Handbook of Security Studies (2nd edition)* (457-468). New York: Routledge.

Dunn-Cavelty, M. & Sutter, M. (2009). Public-Private Partnerships are no silver bullet: An expanded governance model for Critical Infrastructure Protection. *International Journal for Critical Infrastructure Protection*, vol. 2(4), 179-187.

Ehrhart, H.G., Hegemann, H. & Kahl, M. (2014). Towards Security Governance as a critical tool: a conceptual outline. *European Security*, 23(2), 145-162.

Firth-Butterfield, K. (2016). *Lucid AI's Ethics Advisory* Panel. Presentation at the Centre for the Study of Existential Risk, University of Cambridge on 3. February 2016. Accessible at https://www.youtube.com/watch?v=w3-wYGbNZU4

Francois, C. (2015). Robots, War and Society. In *Defense Dossier*, American Policy Council, February 2015 (13), 11-15.

Fukuyama, F. (2016). Governance: What do we know and how do we know it? *Annual Review of Political Science*, 19, 89-105.

Future of Life Institute (2017). *Asilomar AI Principles*. Retrieved from https://futureoflife.org/ai-principles/

Future of Life Institute (2015). *Autonomous Weapons: an Open Letter from AI & Robotics Researchers*. Retrieved from https://futureoflife.org/open-letter-autonomous-weapons/

Gordon, K. and Miyake, M. (2000). Business Approaches to Combating Bribery: A Study of Codes of Conduct, *OECD Working Papers on International Investment*, 2000/01. Paris: OECD Publishing.

Grossman, S.J. & Hart, O. D. (1983). An Analysis of the Principal Agent-Problem. *Econometrica*, 51(1), 7-46.

Gubrud, M. (2016, 1. June). Why Should We Ban Autonomous Weapons? To Survive. *IEEE Spectrum (online)*. Retrieved from http://spectrum.ieee.org/automaton/robotics/military-robots/why-should-we-ban-autonomous-weapons-to-survive

Haas, M. (2014). Autonomous Weapon Systems: The Military's Smartest Toys? *The National Interest (online)*. Retrieved from http://nationalinterest.org/feature/autonomous-weapon-systems-the-militarys-smartest-toys-11708

Hall, R. B. & Biersteker, T. (Eds.) (2002). *The Emergence of Private Authority in International Affairs*. Cambridge: Cambridge University Press.

Haufler, V. (2001). *A Public Role for the Private Sector: Industry Self-Regulation in a Global Economy*. Washington: Carnegie Endowment for International Peace

Hern, A. (2017, 26. January). Whatever happened to the DeepMind AI ethics board Google promised? *The Guardian (online)*. Retrieved from

https://www.theguardian.com/technology/2017/jan/26/google-deepmind-ai-ethics-board

Hernandez, D. & King, R. (2016, 24 November). Universities AI talent poach by tech giants. *Wall Street Journal (online).* Retrieved from https://www.wsj.com/articles/universities-ai-talent-poached-by-tech-giants-1479999601

Horowitz, M. (2016). Who'll want artificially intelligent weapons? ISIS, democracies or autocracies? *Bulletin of the Atomic Scientist (online)*. Retrieved from http://thebulletin.org/who'll-want-artificially-intelligent-weapons-isis-democracies-or-autocracies9692

Hsu, J. (2017, 27 January). Tech Leaders Are Just Now Getting Serious About the Threats of AI. *Backchannel Magazine*. Retrieved from https://backchannel.com/silicon-valleys-ai-ethics-boards-shouldn-t-be-secret-c3f8278d858f#.iz1o6umny

IEEE-SA (2016). *IEEE Standards Association Introduces Global Initiative for Ethical Considerations in the Design of Autonomous Systems.* Retrieved from http://standards.ieee.org/news/2016/ieee_autonomous_systems.html

International Committee of the Red Cross (2015). *Autonomous weapon systems: Is it morally acceptable for a machine to make life and death decisions?* Retrieved from https://www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS

Keohane, R. O. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.

Krahmann, E. (2003). Conceptualizing Security Governance. *Cooperation and Conflict: Journal of the Nordic International Studies Association*, 38(1), 5-26.

Krasner, S.D. (1983). *International Regimes*. Ithaca: Cornell University Press.

Larsen, J. (2002). *Arms Control: Cooperative Security in a Changing Environment*. Boulder/London: Lynne Rienner Publishers.

Lewis, C. (2016, September 6). Capturing Flying Insects: A Machine Learning Approach to Targeting. *War on the Rocks*. Retrieved from http://warontherocks.com/2016/09/capturing-flying-insects-a-machine-learning-approach-to-targeting/

Lincoln Kitman, J. (2016, 19. December). Google Wants Driverless Cars But Do We? *New York Times (online)*. Retrieved from http://www.nytimes.com/2016/12/19/opinion/google-wants-driverless-cars-but-do-we.html?_r=0

Lok, C. (2009, 4. November). Gene-Makers Put forward Biosecurity Standard, *Nature News*. Retrieved from http://www.nature.com/news/2009/091104/full/news.2009.1065.html

Lynn III, W. J. (2014). The End of the Military-Industrial Complex: How the Pentagon Is Adapting to Globalization. *Foreign Affairs*, 93 (6), 104-110.

Marchant, G. E. (2011). The Growing Gap Between Emerging Technologies and the Law. In Marchant, G. E., Allenby, B. R. & Herkert, J. R. (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (19-33). Springer Science and Business.

Markoff, J. (2016, 1. September). How Tech Giants Are Devising Real Ethics for Artificial Intelligence. *New York Times (online)*. Retrieved from https://www.nytimes.com/2016/09/02/technology/artificial-intelligence-ethics.html?_r=0

Matten, D. & Crane, A. (2005). Corporate Citizenship: Toward an Extended Theoretical Conceptualization. *The Academy of Management Review*, 30(1), 166-179.

McCreight, R. (2013). Convergent Technologies and Future Strategic Security Threats. *Strategic Studies Quarterly*, Winter 2013, 11-19.

Mearsheimer, J. (1994). The False Promise of International Institutions, *International Security*, 19(3), (Winter 1994/1995), 5-49.

Moses, L. B. (2007). Recurring Dilemmas: The Law's Race to Keep Up with Technological Change. *University of Illinois Journal of Law, Technology & Policy*, 2007 (21), 239–285.

Nagayoshi, S. (2015). *Statement by Japan*. Retrieved from http://www.unog.ch/80256EDD006B8954/(httpAssets)/7C284472DDBBA998C1257E26005EA5DC/$file/2015_LAWS_MX_Japan.pdf

Nielsen, D. L. & Tierney, M. J. (2003). Delegation to International Organizations: Agency Theory and World Bank Environmental Reform. *International Organizations*, 57 (2), 241-276.

Papadopoulos, Y. (2013). The challenge of transnational private governance: Evaluating authorization, representation, and accountability. *Sciences Po LIEPP Working Paper No 8*. Retrieved from http://www.sciencespo.fr/liepp/sites/sciencespo.fr.liepp/files/WP8.pdf

Parker C, Braithwaite J (2003) Regulation. In Cane P, Tushnet M (Eds.), *The Oxford Handbook of Legal Studies* (119–145). Oxford: Oxford University Press.

Partnership on AI (2016). *Goals*. Retrieved from https://www.partnershiponai.org/#s-goals

Pauewelyn, J., Wessel, R.A. & Wouters, J. (2012). The Stagnation of International Law. *Working Paper No. 97*. Retrieved from http://doc.utwente.nl/86725/1/SSRN-id2271862.pdf

Pitofsky, R. (1998). *Self-Regulation and Anti-Trust*. Statement at DC Bar Association Symposium, February 18, 1998. Retrieved from https://www.ftc.gov/public-statements/1998/02/self-regulation-and-antitrust

Pollack, M. A. (1997). Delegation, agency, and agenda setting in the European Community. *International Organizations*, 51 (1), 99-134.

Ronit, K. (2011). Transnational Corporations and the Regulation of Business at the Global Level. In Reinalda, B. (Ed.). *The Ashgate Companion to Non-State Actors* (75-87). Farnham: Ashgate.

Scherer, A. G., Baumann-Pauly, D. & Schneider, A. (2012). Democratizing Corporate Governance: Compensating for the Democratic Deficit of Corporate Political Activity and Corporate Citizenship. *Business & Society*, 52(3), 473-514.

Stiglitz, J. S. & Wallstein, S. E. (1999). Public-Private Technology Partnerships: Promises and Pitfalls. *American Behavioral Scientist*, 43 (1), 52-73.

Stilgoe, J. & Maynard, A. (2017, 1. February). It's time for some messy, democratic discussions about the future of AI. *The Guardian (online)*. Retrieved from https://www.theguardian.com/science/political-science/2017/feb/01/ai-artificial-intelligence-its-time-for-some-messy-democratic-discussions-about-the-future

Stoker, G. (1998). Governance as theory: five propositions. *International Social Science Journal*, vol. 50 (155), 17-28.

Strowsky, J. (1996). The Dual-Use Dilemma. *Issues in Science and Technology*, vol. XIII (2), Winter 1997.

UN (2012). *First Industry Conference on Security Council Resolution 1540.* Retrieved from https://www.un.org/disarmament/update/20120425/

UN Office Geneva (2017a). *The Convention on Certain Conventional Weapons*. Retrieved from http://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument.

UN Office Geneva (2017b). *2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS).* Retrieved from http://www.unog.ch/80256EE600585943/(httpPages)/F027DAA4966EB9C7C12580CD0039D7B5?OpenDocument

Vogel, D. (2008). Private Global Business Regulation. *Annual Review of Political Science*, 11, 261-282.

Wallach, W. & Allen, C. (2012). Framing Robot Arms Control. *Ethics and Information Technology*, 15, 125–135.

Waltz, K. (1979). *Theory of International Politics*. New York: McGraw Hill.

Wittes, B. & Blum, G. (2015). *The Future of Violence: Robots and Germs, Hackers and Drones—Confronting A New Age of Threat.* New York: Basic Books.

Wolf, D. (2008). Emerging patterns of global governance: the new interplay between the state, business and civil society. In Scherer, A. G. & Palazzo, G. (Eds.) *Handbook of research on Global Citizenship* (225-247). Cheltenham: Edward Elgar.

World Economic Forum (2017). *The Global Risks Report 2017, 12th Edition*. Davos: World Economic Forum. Retrieved from http://www3.weforum.org/docs/GRR17_Report_web.pdf

World Economic Forum (2015). Engineering the Future: How Can the Risks and Rewards of Emerging Technologies Be Balanced? In *Global Risks 2015: 10th Edition*, World Economic Forum. Retrieved from http://www3.weforum.org/docs/WEF_Global_Risks_2015_Report15.pdf

Youngjip, AHN (2015). *Statement by Ambassador AHN Youngjip*. Retrieved from http://www.unog.ch/80256EDD006B8954/(httpAssets)/2A22908A9A03E949C1257E29005B90C1/$file/2015_LAWS_MX_RoK_GS+Corr.pdf