International Review of the Red Cross (2020), **102** (913), 261–285.
Digital technologies and war
doi:10.1017/S1816383121000059

IRRC

The changing role of multilateral forums in regulating armed conflict in the digital age

Amandeep S. Gill
Ambassador Amandeep S. Gill is Director of the International Digital Health and AI Research Collaborative at the Global Health Centre of the Graduate Institute for International and Development Studies in Geneva. He was the Executive Director and Co-Lead of the Secretariat of the UN Secretary-General's High-Level Panel on Digital Cooperation until August 2019 and has served as India's Ambassador and Permanent Representative to the Conference on Disarmament in Geneva. Ambassador Gill chaired the Group of Governmental Experts of the Convention on Certain Conventional Weapons on emerging technologies in the area of lethal autonomous weapon systems from 2017 to 2018. He currently serves as a member of the UNESCO Ad Hoc Expert Group drafting a recommendation on the ethics of artificial intelligence, and as a Commissioner on the *Lancet/Financial Times* Commission on "Governing Health Futures 2030: Growing Up in a Digital World".

Abstract
This article examines a subset of multilateral forums dealing with security problems posed by digital technologies, such as cyber warfare, cyber crime and lethal autonomous weapons systems (LAWS).[1] It identifies structural issues that make it difficult for multilateral forums to discuss fast-moving digital issues and respond in time with the required norms and policy measures. Based on this problem analysis,

and the recent experience of regulating cyber conflict and LAWS through Groups of Governmental Experts, the article proposes a schema for multilateral governance of digital technologies in armed conflict. The schema includes a heuristic for understanding human–machine interaction in order to operationalize accountability with international humanitarian law principles and international law applicable to armed conflict in the digital age. The article concludes with specific suggestions for advancing work in multilateral forums dealing with cyber weapons and lethal autonomy.

Keywords: digital technology, conflict, cyber security, autonomous weapons, human–machine interface, distributed governance, multilateral forums, international humanitarian law, accountability.

Introduction

Global security and stability are increasingly intertwined with digital technologies, and even older-generation cyber capabilities are "becoming more targeted, more impactful on physical systems, and more insidious at undermining societal trust".[2] The challenge has grown beyond the kinetic impact of cyber attacks on critical infrastructure – ports, air traffic control, power grids and financial flows – to the "hacking" of public opinion and political institutions. Efforts to develop international norms for responsible behaviour, build capacity and enhance trust remain fragmented, particularly at the inter-governmental level. Norms that are championed by one side or the other often lack comprehensiveness and a critical mass of support from key governments. Frustrated by a lack of international cooperation, many jurisdictions such as the United States and European Union are increasingly imposing unilateral sanctions on specific individuals and entities in response to cyber attacks.[3] Adding to the mistrust are export control, competition policy and investment policy measures targeted at digital companies of peers.[4]

The advent of artificial intelligence (AI) brings another urgent dimension to the regulation of armed conflict in the digital age. The convergence of three trends – namely, increased computing power, big data sets leveraged through the Internet, and low costs of data storage and manipulation – has mainstreamed AI techniques such as machine learning. AI systems can handle tasks that previously

1 In the author's understanding, digital technologies are devices, platforms, data storage and processing architectures, algorithms, computing languages, communication protocols and standards that rely on the representation of information as discrete binary values. Information and communications technology (ICT) is another term that is interchangeably used in this regard.

2 High-Level Panel on Digital Cooperation, *The Age of Digital Interdependence: Report of the UN Secretary-General's High-Level Panel on Digital Cooperation*, June 2019, p. 27.

3 Patryk Pawlak and Thomas Biersteker (eds), *Guardian of the Galaxy: EU Cyber Sanctions and Norms in Cyberspace*, Chaillot Paper 155, EU Institute for Security Studies, October 2019.

4 Ana Swanson, "U.S. Delivers Another Blow to Huawei with New Tech Restrictions", *New York Times*, 15 May 2020; Li Sikun, "China Ready to Target Apple, Qualcomm, Cisco and Boeing in Retaliation against US' Huawei Ban: Source", *Global Times*, 15 May 2020.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

seemed reserved for human minds. The March 2016 defeat of Lee Sedol, the eighteen-time world champion in the board game go, by Deep Mind's AI algorithm is a powerful example.[5] In particular, the use of machine learning-based AI systems in armed conflict raises concerns about the loss of human control and supervision over the conduct of war.[6] This could result in harm to civilians and combatants in armed conflict in contravention of international humanitarian law (IHL); it could also bring about a new arms race and lower the threshold for the use of force.

Are multilateral forums fit for purpose for the regulation of armed conflict in the digital age? To answer this question, the following analysis first recalls the historical context of modern multilateral forums dealing with conflict prevention and arms control. It thereafter juxtaposes the established procedures and outcomes of such forums against the nature of digital technologies and the unique characteristics of their use in armed conflict. This is followed by a survey of select forums dealing with the consequences of such use, in particular the Group of Governmental Experts (GGE) on Developments in the Field of Information and Telecommunications in the Context of International Security established by the United Nations General Assembly from time to time since 1998 (UN GGE), and the GGE comprised of all High Contracting Parties to the Convention on Certain Conventional Weapons dealing with lethal autonomous weapons systems (LAWS) since 2016. A framework for understanding human accountability under national and international law in the various stages of the development and use of autonomous digital technologies is abstracted from this survey. The analysis concludes with a mapping of some future directions for multilateral forums dealing with cyber conflict and autonomous weapons, and proposes specific steps for multilateral governance of digital technologies in the international security context by co-opting additional governance actors and taking a tiered approach to the application of norms.

Historical background and context

Historically, the idea of international conferences and forums to negotiate the prevention and regulation of conflict *ante factum* in peacetime is relatively new. We can date it back to the post-war conferences in Geneva of 1863–64 and 1868, which codified the rules of war on land and at sea;[7] the St Petersburg Declaration of 1868, which prohibited the use in conflict of a specific category of

5 Amandeep S. Gill, "The Role of the United Nations in Addressing Emerging Technologies in the Area of Lethal Autonomous Weapons Systems", *UN Chronicle*, Vol. 55, No. 3–4, 2019.

6 The UN Secretary-General has called for a ban on machines with the power and discretion to take lives without human involvement. António Guterres, "Remarks at the 'Web Summit'", Lisbon, 5 November 2018, available at: www.un.org/sg/en/content/sg/speeches/2018-11-05/remarks-web-summit (all internet references were accessed in January 2021).

7 The former resulted in the establishment of the International Committee of the Red Cross (ICRC) and the adoption of the Convention for the Amelioration of the Condition of the Wounded in Armies in the Field of 22 August 1864. The latter helped adapt the principles of the Geneva Convention of 1864 to sea warfare.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

munitions;[8] and the series of Peace Conferences held at The Hague in 1899 and 1907. The latter were the "first truly international assemblies meeting in time of peace for the purpose of preserving peace".[9] Their legacy is still with us, for example, in the form of the Permanent Court of Arbitration, which provides a forum for arbitration, enquiry and conciliation among States with regard to the agreements they have entered into, and the famous Martens Clause, which provides a touchstone beyond legal norms for humanitarian protection during armed conflict.[10]

Subsequent forums such as the UN General Assembly and the Military Staff Committee, a subsidiary body of the UN Security Council, reflect the relatively modest provisions of the UN Charter in the field of conflict prevention, disarmament and arms control in comparison, for example, to the Statute of the League of Nations, which was more forward-leaning on collective action for disarmament and world peace.[11] The UN Charter was also pre-atomic, as the provisions of the Charter had been negotiated before the knowledge of atomic weapons became widely known. The development of nuclear weapons and the Cold War arms race further turned attention away from conventional-weapons-related arms control, which tended to play second fiddle to strategic weaponry during the Cold War years and was often seen through a regional rather than a global lens. This gap between the treatment of weapons of mass destruction and conventional weapons persisted even though technology and security trends began to shift in the late 1990s, raising the importance of the latter. The intangible attributes of weapons systems, in particular the growing digitalization of the key components of these systems, also stayed largely unappreciated outside a restricted circle of practitioners of export control regimes such as the Wassenaar Arrangement. The policy-makers' lag with regard to the rapid pace of technology developments in the post-Cold War period is evident also from the shifting use of terms such as "cyber security", "network security", "internet security" and "digital security" – they were unsure of what they were dealing with and tended to echo the most fashionable term of the day.

The digital challenge to multilateral forums

It matters that the "digital turn" was not as dramatic as its nuclear predecessor. It took a while for the practitioners of conflict regulation and arms control to realize that they were dealing with a fundamentally new technology. The internet

8 Declaration Renouncing the Use, in Time of War, of Explosive Projectiles under 400 Grammes Weight, St Petersburg, 29 November and 11 December 1868.

9 James Brown Scott, "Prefatory Note", *The Proceedings of the Hague Peace Conferences: Translation of the Official Texts: The Conference of 1899*, Oxford University Press, Oxford, 1920, p. v.

10 Theodor Meron, "The Martens Clause, Principles of Humanity, and Dictates of Public Conscience", *American Journal of International Law*, Vol. 94, No. 1, 2000.

11 Leland M. Goodrich, Edward Hambro and Anne Patricia Simons, *Charter of the United Nations: Commentary and Documents*, 3rd revised ed., Columbia University Press, New York, 1969.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

protocols had been known to the esoteric few since the 1970s, but the switch to the common Transmission Control Protocol/Internet Protocol (TCP/IP) for all internet hosts in 1984 and the invention of the World Wide Web at CERN in Geneva in 1989 led rapidly to the global adoption of digital communications networks. A vast array of applications began to be deployed on the virtual infrastructure created by this distributed and scalable plumbing. The marriage of internet and telephony at the beginning of the twenty-first century liberated users from the yoke of a desk, and social media platforms allowed them to create content at an unprecedented scale. That included, of course, spam, incitement to hate and violence, and malware. Digital networks permitted an unprecedented level of anonymity in the generation and use of content, the Internet's global reach permitted rapid virality of such content, and the high stakes involved in access to and use of the Internet created the incentives for malicious use at scale.

Why is this important in the context of IHL and arms control? Force has traditionally been correlated with physical destruction, injury and death, and few had imagined that non-physical destruction or the physical consequences of non-physical attacks could be so significant as to become issues of international security. Further, if misinformation and propaganda had needed to be regulated, it had been in the narrow context of the ruse in combat or the misuse of symbols of neutrality and protection. No one had imagined that disinformation and distortion of ideas and institutions through digital technology could reach such scale. Other certainties – combatant versus non-combatant, civilian versus military target – were also shaken, if not compromised. Finally, even though irregulars have been used throughout the history of IHL and arms control, practitioners had always assumed that attribution for the use of force was a tractable problem. How to assign accountability when there are so many fingers on the digital trigger, many unwittingly so, and when arguably there might be no trigger at all?

Having examined the conceptual challenges, let us look at the challenges around practice. Traditionally, multilateral forums have regulated conflict by helping to negotiate norms that regulate, restrict or rule out specific means and methods of warfare, as well as by promoting dialogue and confidence-building among potential belligerents. One category of forums – though admittedly small – has allowed States to clarify ambiguous situations, detect cheating and mediate differences before they snowball into conflict.[12] The toolkit that such forums have deployed in architecting norms include measures for information exchange and transparency, declaration of facilities and stocks, measures for verifying the absence of certain types of activities and items, and compliance measures for punishing violations.[13] In technology areas such as space, chemicals, biology and nuclear science, where both civilian and military uses exist, multilateral regimes

12 An example is the International Atomic Energy Agency (IAEA) investigation of ambiguities related to Iran's uranium enrichment centrifuge programme. IAEA, "Verification and Monitoring in Iran", available at: www.iaea.org/newscenter/focus/iran.

13 The Open Skies Treaty of 1992, for example, uses aerial inspections to verify deployment of forces in order to rule out surprise attacks. Arms Control Association, "The Open Skies Treaty at a Glance", fact sheet, available at: www.armscontrol.org/factsheets/openskies.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

such as the Missile Technology Control Regime regulate transfers through lists of sensitive items and guidelines for exports to prevent diversion of civilian technologies for military use.[14]

However, digital technologies pose unique challenges for the practitioners of such forums.[15] The question of what exactly constitutes a tool of conflict does not have an obvious answer, and applying the traditional lens of dual-use technologies – space launch vehicles versus missiles, nuclear power reactors versus fissile material production reactors – does not result in an actionable insight. Unlike a battle tank or a ballistic missile, no digital tool can be deemed a weapon independent of its context.[16] Such tools can be endlessly replicated, repurposed and combined with physical systems. There is no discrete set of "processes" such as the two routes of enrichment and reprocessing for producing fissile material for nuclear weapons, and there are no easily identifiable "choke points", such as separation of plutonium from spent fuel, which can be policed.[17] In terms of effect, it is often hard to establish the threshold at which something jumps from the virtual to the physical, from surveillance and harassment to inter-State conflict, and from the local and the national into the international domain. Unlike physical weapons, cyber weapons do not destroy themselves and can be reused to attack the attacker.[18] The actors, too, are hard to separate into discrete sets of State and non-State actors, as States seldom claim official responsibility for cyber attacks. Furthermore, the category of State actors is itself fluid – there is no privileged and static group of possessors behind a high entry threshold.[19] Notions of parity and balance so dear to Cold War practitioners of arms control are also hard to define in matters digital.[20]

At a more basic level, the speed and spread of development of digital technologies overwhelms the ability of policy forums to keep pace with the social, economic and political consequences of technological change.[21] This is further complicated by the bigger role (compared to the State sector) that the private

14 See the "MTCR Guidelines" and "MTCR Annex" sections of the Missile Technology Control Regime website, available at: https://mtcr.info.

15 Colin Picker, "A View from 40,000 Feet: International Law and the Invisible Hand of Technology", *Cardozo Law Review*, Vol. 23, 2001, p. 149.

16 Take malware, for example: law enforcement agencies can use the same lines of code to monitor and thwart the planning of terrorist acts.

17 Funding of research and development (R&D), certain types of datasets or computing capacity could theoretically be considered choke points for the development of lethal autonomous weapons, but practical ways to prevent their use for military purposes are impossible to envision.

18 An example is the NotPetya malware, which used a penetration capability, named EternalBlue, allegedly developed by the US National Security Agency but leaked in early 2017. David Bisson, "NotPetya: Timeline of a Ransomworm", *Tripwire*, 28 June 2017.

19 Unlike the Treaty on the Non-Proliferation of Nuclear Weapons, for example, which designates five States that manufactured and exploded a nuclear device prior to 1967 as nuclear weapon States (Art. IX, para. 3).

20 In the case of nuclear weapons and their delivery systems, elaborate models of parity, stability and balance can be built because their impact is knowable and their capabilities can be estimated, counted and even enshrined in treaty systems such as the Soviet–US strategic treaties. However, cyber capabilities are highly esoteric. Their impact independent of context is hard to estimate, let alone compare.

21 The so-called "Moore's law", which states that density of transistors on an electronic chip will double every year even as costs go down, is an illustration of this frenetic pace.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

sector now plays in the development of cutting-edge technology, business applications and the creation of intellectual property.[22] Table 1 summarizes some of the characteristics of traditional multilateral forums which make it challenging to address the policy implications of digital technologies.

Table 1. *Attributes of multilateral forums which pose challenges with regard to digital issues*

Attribute	Challenge
Periodicity, response time	Multilateral forums meet at regular intervals, often annually, even though working groups and preparatory committees may meet inter-sessionally. Meetings last for a short period of time, ranging from less than a week to a few weeks. Treaty negotiation takes many years and treaty review conferences often happen every five years. At the digital end, one year is a long time in technology development and adoption, and irregular meetings for a short duration are insufficient to study impact and plan policy responses.[23]
Agenda-setting	Multilateral forums have structured agendas and inter-sessional programmes with negotiated mandates. Technology issues are often tucked under broad items such as "science and technology developments" or the "impact of emerging technologies". Digital technologies are hard to fit into static agendas and mandates; as the Cambridge Analytica case highlights, they require specificity with regard to the context of their use for governance issues to clearly emerge.[24]

Continued

22 The 2019 *Digital Economy Report* by the UN Conference on Trade and Development (UNCTAD) has an excellent analysis of the policy challenges posed to traditional forums on competition policy and trade rules by the explosive growth of the digital economy. UNCTAD, *Digital Economy Report 2019*, Geneva, 2019, available at: https://unctad.org/en/PublicationsLibrary/der2019_en.pdf.

23 Mark Zuckerberg's testimony before the US Congress is an example of the lag not only between policy-making and digital technology's impact but also between policy-makers' understanding of complex and fast-moving technology-based business models and those who create and run such businesses. Casey Newton, "The 5 Biggest Takeaways from Mark Zuckerberg's Appearance before the Senate", *The Verge*, 10 April 2018. available at: www.theverge.com/2018/4/10/17222444/mark-zuckerberg-senate-hearing-highlights-cambridge-analytica.

24 *Ibid.*

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

Table 1. *Continued*

Attribute	Challenge
State-centrism	Multilateral forums are State-centric, and State participation is usually limited to diplomatic representatives. Inter-ministerial coordination prior to and inter-agency participation in multilateral meetings is often a luxury for most delegations.[25] Digital technologies are developed largely in the private sector, and this has implications for both understanding and control of technology by governments. Their regulation also generally falls under several ministries and departments, making coordinated action difficult.[26]
Tangibility	Multilateral forums dealing with arms control and conflict focus on tangible weapons and tools, specific production pathways, and clearly delineated industry domains and business models. The impact of these artefacts is also seen in tangible terms – destruction of property, loss of life, damage to the environment etc. Digital technologies are intangible; they may not be part of weapons but can still influence conflict significantly (e.g. through decision support systems) and cannot be isolated from other technology domains. Unlike other weapons, the impact of digital technologies can be mostly (though not necessarily exclusively) socio-psychological, without an underlying physically destructive threat. This attribute can disorient practitioners who are used to counting, comparing and controlling discrete weapons and platforms.
Outcomes	The output of these forums is either legally binding treaties, "hard law" that can be implemented by governments through domestic legislation and/or regulations, or reports and resolutions that have a political impact in the context of inter-State relations. Treaties are policed by verification regimes with inspections, declarations and challenge inspections.

25 "Small Developing Countries Struggle in WTO", *Forbes,* 19 May 2010.
26 For example, unlike traditional banking, which is regulated through ministries of finance, digital payments cut across finance, communications and ICT ministries.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

	Political commitments are upheld through dialogue, diplomacy, reciprocity and reputational consequences. Hard law applies to only a narrow aspect of digital technologies, chiefly business-related regulation, consumer and worker protections etc. Mostly, however, digital technologies are governed through "soft law", including "voluntary programs, standards, codes of conduct, best practices, certification programs, guidelines, and statements of principles".[27] Intrusive verification is not standard practice for enforcing these norms.
Inter-disciplinarity	Multilateral forums, by design and in their functioning, tend towards specialized treatment of subjects (trade, disarmament, human rights, environment etc.) independent of technology. While this is changing,[28] most participants in these forums, and the secretariats supporting them, also tend to come from a specialized field. The lack of cross-disciplinary approaches is particularly striking in arms control and disarmament forums, where practitioners normally do not interact with technologists and entrepreneurs.[29] For their part, technology developers and entrepreneurs often lack appreciation of the political and security impact of their innovations. Opportunities for engagement with policy-makers are limited to market regulatory contexts, although in recent years multi-stakeholder forums such as the World Economic Forum have expanded the circle.
Power distribution	Multilateral forums operate with traditional power dynamics – major powers often act as "norm

Continued

27 Wendell Wallach and Gary Marchant, "Toward the Agile and Comprehensive International Governance of AI and Robotics", *Proceedings of the IEEE*, Vol. 107, No. 3, 2019.

28 Recent initiatives on internet governance and digital cooperation have used digital technologies and cross-disciplinarity of impact as their organizing principles, rather than a specific UN domain such as human rights.

29 In the author's experience of negotiations at the Conference on Disarmament or discussions at the UN Disarmament Commission from 2010 to 2017, there was not a single instance of interaction with industry; this is unlike the Nuclear Security Summit process, which was organized outside the UN context. In the Convention on Certain Conventional Weapons (CCW) context, industry interaction was brought in through side events in 2017–18.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

Table 1. *Continued*

Attribute	Challenge
	entrepreneurs",[30] while groupings of States act as pressure groups. While private companies and civil society have had an important agenda-setting and opinion-shaping role in some discussions,[31] they take a secondary position to more powerful State and inter-State actors. This power asymmetry sits uneasily with the digital technology reality. For example, digital platforms such as Facebook, Alipay and WhatsApp may have more users ("virtual residents") than the populations of most countries; they operate quasi-global infrastructures, act as cross-border "content policemen" and have market capitalizations that dwarf other sectors and most national GDPs. If norms related to digital technologies are to have an impact, the digital industry has to be a part of the discussion on policy responses and has to cooperate with State actors for their implementation.

Against the background of these challenges, it is useful to look at select multilateral forums at the interface of digital technologies and international security.

Select multilateral forums dealing with international security implications of digital technologies: The UN GGEs and the OEWG on information security

Upon an initiative of the Russian Federation, the item "Developments in the field of information and telecommunications in the context of international security" was put on the agenda of the UN General Assembly in 1998. Resolutions adopted by the First Committee of the General Assembly under this item have over the years created five GGEs to examine the issue and make recommendations. A sixth one is currently under way under the chairmanship of Brazil's Guilherme Patriota, alongside an Open-Ended Working Group (OEWG) on the same subject under

30 Martha Finnemore and Kathryn Sikkink, "International Norm Dynamics and Political Change", *International Organization*, Vol. 52, No. 4, 1998.

31 Two examples are the 1997 Anti-Personnel Landmines Convention, or Ottawa Treaty, and the 2017 Treaty on the Prohibition of Nuclear Weapons. Civil society's role in achieving these treaties was acknowledged through the award of the Nobel Peace Prize to Jody Williams and the International Campaign to Ban Landmines in 1997, and to the International Campaign to Abolish Nuclear Weapons in 2017.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

the chairmanship of the Swiss ambassador Jürg Lauber.[32] The limited-membership GGEs, with fifteen to twenty-five government-appointed experts, hold their sessions alternately in Geneva and New York, while the OEWG is open to all UN member States and observers and meets in New York. The value of the UN GGE and OEWG processes is in sustaining dialogue to develop common understandings on the security implications of the use of digital technologies as well as in promoting norms, rules and principles for responsible State behaviour in this area. Opportunities for industry, academia and civil society to engage with these forums, albeit in a limited fashion (such as through written inputs or short statements during plenary debates), have been provided for as part of the resolutions setting them up.[33]

The report of the 2013 UN GGE affirmed that international law, and in particular the UN Charter, is applicable and essential to maintaining peace and stability and promoting an open, secure, peaceful and accessible information and communications technologies (ICT) environment. The 2015 GGE further proposed a set of eleven voluntary and non-binding norms for States aimed at promoting an open, secure, stable, accessible and peaceful ICT environment, while noting a proposal for a Code of Conduct on Information Security.[34] These norms include an obligation (voluntary) not to conduct or knowingly support ICT activity contrary to a State's obligations under international law that intentionally damages critical infrastructure or otherwise impairs the use and operation of critical infrastructure to provide services to the people.[35] They also include an injunction not to target the computer emergency response services of other States. This recalls the notion of protected objects/functions and the principle of distinction in IHL. Unfortunately, consensus broke down in the 2017 GGE, chiefly on the issue of the applicability of international law to cyber conflict.[36]

The challenge today is not only repairing that breach in consensus but also reconciling the outputs of two forums moving in parallel pathways on the same issue: the sixth UN GGE and the OEWG, which were set up in December 2018 through resolutions championed by the United States and Russia respectively. The fault line is not only procedural but also substantive: it lies between approaches that stress the applicability to cyber conflict of existing norms and accountability thereunder, on the one hand, and approaches that reject an

32 The new GGE was set up upon an initiative of the US, while Russia, a previous votary of the GGE idea, switched in 2018 to sponsoring the more participatory methodology of an OEWG open to all member States while still participating in the limited-membership GGE set up under US sponsorship.

33 Details of multi-stakeholder engagement and inputs with regard to the OEWG are available on the OEWG website, available at: www.un.org/disarmament/open-ended-working-group/.

34 UN GGE, *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, UN Doc. A/70/174, 22 July 2015, paras 12, 13.

35 Recent multi-stakeholder processes such as the Global Commission on the Stability of Cyberspace (GCSC) have suggested additional norms and protections, including for electoral infrastructure. GCSC, *Advancing Cyberstability: Final Report*, November 2019.

36 Elaine Korzak, "UN GGE on Cybersecurity: The End of an Era?", *The Diplomat*, 31 July 2017, available at: https://thediplomat.com/2017/07/un-gge-on-cybersecurity-have-china-and-russia-just-made-cyberspace-less-safe/.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

automatic extension of existing international law, in particular IHL, to cyber conflict and stress instead the need to negotiate new norms, on the other.

It is instructive in this regard to juxtapose the comments of the Russian and US delegates on the pre-draft of the OEWG's report circulated by the chair on 11 March 2020.[37] The US delegate, in comments largely appreciative of the chair's pre-draft but critical of certain inputs, stated:

> We appreciate that the draft report memorializes that all States reaffirmed that international law and, in particular, the Charter of the United Nations, apply to the use of ICTs by States. … The present draft devotes far too much attention (paragraphs 27–30) to proposals made by a minority of States for the progressive development of international law, including through the development of a legally binding instrument on the use of ICTs by States. These proposals lacked specificity and are impractical. The OEWG's mandate is to study how international law applies to the use of ICTs by States, and the report should therefore focus on existing international law. Without a clear understanding of States' views on how existing international law applies to ICTs, it is premature to suggest that international law needs to be changed or developed further.[38]

The Russian delegate, while noting some "positive traits", referred to "many unacceptable approaches" and stated:

> The document exaggeratedly emphasizes the principle of applicability of universally recognized norms and principles of international law set forth in the UN Charter and in the Declaration on principles of international law concerning friendly relations and cooperation among States in accordance with the Charter of the United Nations, dated 1970, to the use of ICTs. At the same time, this principle is not linked to specific modalities of its applicability, namely, who, how and in which circumstances can apply it. These practical aspects demand to be regulated by a specialized international legal instrument that would provide for the modalities of applicability of the existing norms of international law to the use of ICTs, as well as, if necessary, includ[ing] new norms. … We regard as potentially dangerous the attempts to impose the principle of full and automatic applicability of IHL to the ICT environment in peacetime. This statement itself is illogical and contradictory because IHL is applied only in the context of a military conflict while currently the ICTs do not fit the definition of a weapon.[39]

This is not the only arena in which these opposite views on existing law versus new norms play out. In the context of cyber crime, Russia, China and others have

37 OEWG, "Initial 'Pre-draft' of the Report of the OEWG on Developments in the Field of Information and Telecommunications in the Context of International Security", 11 March 2020, available at: https://unoda-web.s3.amazonaws.com/wp-content/uploads/2020/03/200311-Pre-Draft-OEWG-ICT.pdf.

38 "United States Comments on the Chair's Pre-draft of the Report of the UN Open Ended Working Group", available at: https://tinyurl.com/yyus5uv7.

39 "Commentary of the Russian Federation on the Initial 'Pre-draft' of the Final Report of the United Nations Open-Ended Working Group", available at: https://tinyurl.com/yxfuudvd.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

proposed through the UN General Assembly the negotiation of a comprehensive international convention on countering the use of information and communications technologies for criminal purposes.[40] Many Western countries see this as potentially inimical to human rights and fundamental freedoms, as well as superfluous given the 2001 Budapest Convention on Cybercrime, which was negotiated within the Council of Europe framework; others see it as a crucial step going forward.[41]

The current geopolitical context of discussions on information security in New York makes it difficult to make progress, within the UN setting, on what constitutes a cyber weapon or cyber attack in the context of existing law on the use of force, how existing norms regulating the use of force would apply, and what gaps, if any, need to be addressed through additional norms. Progress is more feasible in smaller settings (or "minilaterals") on what norms apply to cyber conflict and cyber crime, and how those norms apply; attributing responsibility for cyber attacks; building mutual confidence and capacity; and promoting cooperation on enforcement of norms on cyber conflict.[42]

The importance of regional efforts is recognized in the resolution that set up the ongoing UN GGE in 2018, which requires the UN Office of Disarmament Affairs

> to collaborate with relevant regional organizations, such as the African Union, the European Union, the Organization of American States, the Organization for Security and Cooperation in Europe and the Regional Forum of the Association of Southeast Asian Nations, to convene a series of consultations to share views on the issues within the mandate of the group in advance of its sessions.[43]

The rival resolution under the same agenda item setting up the ongoing OEWG similarly calls upon the UN to encourage regional efforts, promote confidence-building and transparency measures, and support capacity-building and the dissemination of best practices.[44]

40 UNGA Res. 74/247, "Countering the Use of Information and Communications Technologies for Criminal Purposes", Draft Resolution, UN Doc. A/C.3/74/L.11, 11 October 2019.

41 "UN Approves Russian-Sponsored, China-Backed Bid on New Cybercrime Convention", *South China Morning Post*, 28 December 2019, available at: www.scmp.com/news/world/united-states-canada/article/3043763/un-approves-russian-sponsored-china-backed-bid-new.

42 For example, a process facilitated by the NATO Cooperative Cyber Defence Centre of Excellence resulted in a compilation of annotated international norms in 2013 known as the Tallinn Manual (updated in 2017 to the Tallinn Manual 2.0). The Tallinn Manual's drafting reflects the view that pre-cyber-era international law applies to cyber operations, both conducted by and directed against States, and that States both have rights and bear obligations under international law. See Michael N. Schmitt and Liis Vihul (eds), *Tallinn Manual 2.0 on International Law Applicable to Cyber Operations*, 2nd ed., Cambridge University Press, Cambridge, 2017. Another example is the proposal by some members of the Shanghai Cooperation Organization for a draft Code of Conduct on Information Security: see "Letter dated 9 January 2015 from the Permanent Representatives of China, Kazakhstan, Kyrgyzstan, Russian Federation, Tajikistan and Uzbekistan", UN Doc. A/69/723, 13 January 2015.

43 UNGA Res. 73/266, "Advancing Responsible State Behaviour in Cyberspace in the Context of International Security", Draft Resolution, UN Doc. A/C.1/73/L.37, 18 October 2018, op. para. 4.

44 UNGA Res. 73/27, "Developments in the Field of Information and Telecommunications in the Context of International Security", Draft Resolution, UN Doc. A/C.1/73/L.27/Rev.1, 29 October 2018, preambular para. 11.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

What is less clear is to what extent UN processes can collaborate with the private sector, which is often the first responder and an involuntary accomplice in cyber attacks, and which is keen to clarify the ambiguity on applying norms in its own interest.[45] Engagement of the private sector could also be essential to avoiding unintended escalation because of a private sector-led response to cyber attacks from another jurisdiction.[46] However, UN forums, for good reasons, cannot treat private sector representatives on a par with member State representatives and often exclude private companies completely. Impatient with progress in multilateral forums, voluntary non-governmental initiatives such as the Paris Call and the Global Commission on the Stability of Cyberspace (GCSC) have taken steps towards crafting common denominators among multiple stakeholders on acceptable behaviour. Such initiatives can complement the efforts of multilateral intergovernmental forums.[47] In recent years, the UN Secretariat and UN agencies have also taken the initiative on involving the private sector more on issues related to technology. The International Telecommunications Union's (ITU) annual AI for Good Summit and the UN Secretary-General's High-level Panel on Digital Cooperation are two examples.

To sum up, the New York-centric efforts on cyber conflict have taken a more purposeful and welcome turn in the past couple of years despite the trying geopolitical circumstances. Challenges remain with regard to the engagement of non-traditional norm shapers, the absence of bottom-up possibilities for norm-making, and the need to depoliticize relatively less controversial aspects of cyber conflict such as assessment, assistance and cooperation.

In Geneva, the focus has traditionally been on expert-driven, in-depth work, aiming to result in legally binding norms on disarmament, arms control, human rights law and IHL. The city, which hosts the "single" multilateral disarmament negotiating forum, the Conference on Disarmament,[48] is also the home of the International Committee of the Red Cross (ICRC) and several other forums, including the World Intellectual Property Organization and the ITU, which plays an important role in the development of standards on digital technologies and capacity-building on cyber security. Geneva's perceived neutrality is an important consideration for multilateral and multi-stakeholder efforts on cyber security; it is therefore not surprising that the UN Institute for Disarmament Research (UNIDIR) holds its annual conference on stability of

45 Tech companies handle millions of attempts to breach cyber security measures on a daily basis. Their servers can be unwitting hosts for distributing malware, and the costs of breaches or non-payment of insurance claims due to ambiguity about the source of attacks can be crippling.

46 The French and Indian experts on the 2017 GGE made proposals to this effect. Source: personal communication with the author.

47 The Final Report of the GCSC proposes, for example, a set of eight norms additional to those proposed by the GGE in 2015, with the proposed norms applying both to State and non-State actors. GCSC, above note 35.

48 The UN General Assembly's First Special Session on Disarmament in 1978 created a "triad" of disarmament forums: the First Committee of the UN General Assembly, in New York; the UN Disarmament Commission, again in New York, as a universal deliberative body; and the Conference on Disarmament, in Geneva, as the "single" multilateral disarmament negotiating forum.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

cyber space in Geneva and that the CyberPeace Institute, with its objectives of assisting victims of cyber attacks and establishing accountability for such attacks, has chosen Geneva as its host city.[49] The extensive ecosystem of humanitarian and human rights mechanisms as well as trade and development institutions in Geneva is an important asset for international cooperation on norms in the digital field, which does not respect traditional boundaries across the three UN pillars of trade and development, peace and security, and human rights and humanitarian affairs.

In a 2014 paper, Joseph Nye described in detail a "regime complex for managing global cyber activities", which left out what to many was an obscure forum in Geneva at the juncture of IHL and arms control.[50] This is 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 (Convention on Certain Conventional Weapons, CCW). The CCW, negotiated under UN auspices in 1979–80, has its roots in key IHL principles such as proportionality and distinction between civilians and combatants. Currently, the Convention has five Protocols – Protocol I on Non-Detectable Fragments, Protocol II on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (as amended on 3 May 1996), Protocol III on Prohibitions or Restrictions on the Use of Incendiary Weapons, Protocol IV on Blinding Laser Weapons, and Protocol V on Explosive Remnants of War, which deals with the problem of unexploded and abandoned munitions. The modular design of the Convention allows new instruments to be attached to the framework treaty as technology evolves and new humanitarian concerns emerge.[51]

It would have been hard to imagine in early 2014 that the CCW, which had hitherto mostly dealt with almost nineteenth-century-type weapons systems, would become a front-line forum dealing with the international security and international law implications of emerging digital technologies.[52] The focus, with incidents such as the Stuxnet attack on Iran's uranium enrichment gas centrifuges, had been on malware and cyber conflict. A series of breakthroughs in machine learning in the 2010s propelled another set of digital technologies loosely known as AI to the forefront, and the CCW became the forum for dealing with AI-based lethal autonomous weapons systems.

Regulating lethal autonomy in weapons systems: The CCW case

An important foundation for the CCW discussion was the 2013 report of the UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof

49 Information on the CyberPeace Institute is available at: https://cyberpeaceinstitute.org/.
50 Joseph S. Nye Jr., *The Regime Complex for Managing Global Cyber Activities*, Global Commission on Internet Governance Paper Series No. 1, May 2014.
51 For details of the CCW and its Protocols, see UN Geneva, "The Convention on Certain Conventional Weapons", available at: https://tinyurl.com/y4orq8q5.
52 A. S. Gill, above note 5.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

Steyns. Special Rapporteurs are independent human rights experts with a considerable degree of autonomy in pursuing their mandates and in seeking evidence and inputs from governments, civil society, academia and industry.[53] Steyns was concerned by the arbitrariness involved in using drones to target non-State actors and how that challenge could be compounded by the use of autonomous technologies. He defined lethal autonomous robotics as "weapon systems that, once activated, can select and engage targets without further human intervention", and underlined concerns about the accountability of such systems and their compatibility with the requirements of IHL and the standards protecting life under international human rights law.[54] The report generated considerable debate but since the subject was felt to be more a question of arms control and the laws of armed conflict than one of human rights, key delegations in Geneva pushed for the consideration to be moved to the CCW in 2014.[55] A series of informal Meetings of Experts from 2014 to 2016 helped build consensus on a mandate in December 2016 for a formal GGE. The CCW GGE is the expert subsidiary body of the CCW, not just for examining issues like the GGEs set up by the First Committee in New York but also for negotiating new protocols when there is agreement to do so.[56] The CCW GGE, along with other GGEs established under the CCW, is thus different in nature and in terms of participation from the GGEs set up by the First Committee in that it is open to all High Contracting Parties to the Convention. In December 2016, a new GGE established under the CCW, the CCW GGE on LAWS, was mandated by the High Contracting Parties to examine "emerging technologies related to lethal autonomous weapons systems".[57]

There was now a forum with a mandate on the issue and with some attributes that mitigated the challenges previously listed. Its modular framework and past practice offered the possibility of graduating from a discussion to the negotiation of a binding instrument. Note that this is not the case with the GGEs established by the First Committee, although an OEWG can shift gears from discussion to negotiation with a fresh mandate.[58] Countries with emerging capabilities in AI systems such as Australia, Brazil, Canada, China, France, Germany, India, Israel, Japan, the Republic of Korea, Russia, South Africa, the United Kingdom and the United States are all party to the CCW. Further, the balance between humanitarian principles and military necessity inherent to

53 UN Office of the High Commissioner for Human Rights, "Special Procedures of the Human Rights Council", available at: www.ohchr.org/en/HRBodies/SP/Pages/Welcomepage.aspx.

54 Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc. A/HRC/23/47, 9 April 2013.

55 Unsurprisingly, these included the main possessors and users of armed drones. Source: personal communication with the author.

56 See the reference to the negotiating work of the 2011 GGE on Cluster Munitions in UN Geneva, "GGE Sessions in 2011", available at: https://bit.ly/2ZozQMI.

57 Decision I of the Sixth Review Conference of the High Contracting Parties to the CCW, 12–16 December 2016.

58 This was the case most prominently with the Arms Trade Treaty adopted in 2013 after a seven-year process. Michael Spies, "Towards a Negotiating Mandate for an Arms Trade Treaty", *Disarmament Diplomacy*, No. 91, Summer 2009.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

the Convention and the principal tenets of IHL provided the space for States with very differing views to begin engaging on a complex and rapidly evolving technology.

This is not to say that the choice of the forum or its continued use for addressing lethal autonomy is trouble-free. Its periodicity is annual, and the time allocated to discussions has varied from one to two weeks per annum.[59] Despite rules allowing the participation of academia, civil society and humanitarian organizations, the CCW is still State-centric. Therefore, direct engagement of AI technologists and industry in rule-making is not possible. Further, while anchoring the discussion in IHL comforts key stakeholders, it puts others who look at the issue essentially from a human rights perspective at unease.[60]

The challenge of interdisciplinarity has been mitigated to some extent with side events held by the ICRC, NGOs, academia and governments, which helped bring in fresh perspectives, including those of entrepreneurs.[61] Discussions in 2017 were focused by design on raising awareness of the technology, sifting through the hype and the dystopian fantasies, teasing out the cross-domain connections, bringing together legal, ethical, military, technical and entrepreneurial perspectives, and moving beyond binary mindsets of civilian-military objects and dual-use technologies.[62] The ethics panel was particularly useful in engaging the human rights and faith-based communities.[63] With regard to tangibility, testimonies and reports by independent experts, think tanks and organizations such as the ICRC, UNIDIR and SIPRI, as well as working papers submitted by the more active national delegations, helped build awareness of the nature of autonomous technologies. The ITU's AI for Good summits, which coincidentally started in 2017 across the road from the Palais des Nations, engaged many national delegations and helped cross-pollinate governance thinking around AI.[64]

59 In 2017, the CCW GGE on LAWS met for a week of five working days; this went up to ten working days over two weeks in 2018 and then came down to seven working days in 2019. The 2020 meetings are back to two weeks, while the time allotted in 2021 could go up to four weeks. The time for the meetings is negotiated as part of the annual mandate for the GGE and is subject to budgetary and political considerations. 2017 was curtailed because of arrears in payments by High Contracting Parties. See meeting recordings on the UN Digital Recordings Portal, available at: https://conf.unog.ch/digitalrecordings/; CCW, *Meeting of the High Contracting Parties to 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: Final Report*, UN Doc. CCW/MSP/2019/9, 13 December 2019.

60 This was the position of many delegations such as those of Sierra Leone, Costa Rica, Mexico, the Holy See and Honduras. For example, Ambassador Yvette Stevens of Sierra Leone argued at the CCW GGE on LAWS session of 13–17 November 2017 that the Human Rights Council should remain seized of the matter in parallel.

61 See, for example, the 2017 schedule of side events of the CCW GGE on LAWS.

62 CCW GGE, *Food-for-Thought Paper*, UN Doc. CCW/GGE.1/2017/WP.1, 4 September 2017. See also CCW GGE, *Provisional Programme of Work*, UN Doc. CCW/GGE.1/2017/2, 4 September 2017, with the integration of cross-disciplinary inputs into the discussions through four expert panels.

63 See meeting records on the UN Digital Recordings Portal, available at: https://conf.unog.ch/digitalrecordings/.

64 Se the AI for Good Global Summit Reports for the years 2017–19, available at: https://aiforgood.itu.int/reports/.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

What about outcomes? In 2017, the CCW GGE on LAWS adopted a set of conclusions and recommendations by consensus.[65] One of these was that the CCW is the appropriate framework for dealing with the issue, while another was that IHL applies fully to the potential development and use of LAWS. This was an important early assurance, although it did not settle the question of whether further legal norms were needed. The consensus conclusions also highlighted three issues for future work: characterization of the systems under consideration – the so-called definitional issue; aspects of human–machine interaction, critical to the concern about potential violations of IHL; and possible options for addressing the humanitarian and international security consequences of such systems. Divergent views persisted on the definitions, risks and possible benefits of LAWS – as well as approaches to regulation and control, including the idea of a pre-emptive ban – but the chair's summary ended up becoming a practical device for capturing this diversity of views on future outcomes without blocking progress on substance.[66]

It is worth comparing the relative ease with which applicability of IHL was accepted in the LAWS context compared with the continued difficulty in the context of cyber weapons in the UN GGE and the OEWG. In the CCW, as was said in response to concerns expressed about the IHL reference by the Chinese delegate in 2017, the context was that of armed conflict and lethality – even if some delegations and the ICRC argued for a broader approach of "autonomous weapons systems".[67] Thus, in the context of the objectives and purposes of the CCW, IHL was clearly relevant regardless of views on whether further clarifications on the application of IHL principles to LAWS and/or new norms were needed or not.

A key concept in discussions on LAWS in 2017 was that of meaningful human control.[68] Conceptually attractive because it provided a way to avoid the negotiation of additional norms for ensuring compliance with IHL, it was for that reason also seen as problematic and even otherwise subject to different interpretations. At the April 2018 meeting of the CCW GGE on LAWS, it became possible to look at the broader notion of human involvement and intervention from the perspective of different parts of the technology development cycle. This allowed the GGE to move beyond the conceptual discussion on "meaningful human control" or similar concepts like "appropriate human judgement" and to look at the quality and depth of the human–machine interaction essential for meaningful compliance with IHL in each phase of technology development, deployment and use. Equally, it allowed governance

65 CCW GGE on LAWS, *Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems*, UN Doc. CCW/GGE.1/2017/CRP.1, 20 November 2017.

66 A. S. Gill, above note 5.

67 See meeting records for 17 November 2017 on the UN Digital Recordings Portal, available at: https://conf.unog.ch/digitalrecordings/.

68 CCW GGE, *Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, in the Context of the Objectives and Purposes of the Convention: Submitted by the Netherlands*, UN Doc. CCW/GGE.1/2017/WP.2, 9 October 2017.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

choices at various levels for ensuring human responsibility and accountability to become clearer. The "sunrise slide" shown in Figure 1 captures this discussion and points the way to a distributed technology governance scheme that integrates multilateral norms with national regulatory approaches and industry self-regulation.[69]

We have previously seen how the absence of an agreed definition of cyber weapons is a barrier to progress on legal norms. The issue of defining LAWS could have been a showstopper in the CCW, but it was set aside for a step-by-step examination. At its April 2018 meeting, the CCW GGE on LAWS used a "LEGO exercise" to stack up different attributes that would be needed to characterize LAWS without picking specific attributes early or focusing only on technical characteristics. For one, an early definitional approach could have left out potential pathways to such weapons systems while the technology is still under development.[70] Thus, common ground was built up on the concepts and characteristics that would be required for an eventual definition without searching for the perfect dividing line between what is autonomous (future) and what is automated (present). Additionally, an understanding was reached that there needed to be a continued review of technology relevant to LAWS both for understanding what it is that delegations were dealing with but also its potential impact on regional and international security.[71]

An important substantive outcome reached in August 2018 was a set of "Possible Guiding Principles".[72] This negotiated outcome would not have been possible without the consensus reached the previous year on paragraph 16 of the CCW GGE on LAWS' November 2017 report, which noted the cardinal principle that IHL continues to apply fully to all weapons systems, including the potential development and use of LAWS.[73] That the Guiding Principles were not the end point but an essential foundation for further work was underlined in the consensus 2018 report, which included four options for a policy response: a legally binding instrument with prohibitions and regulations on LAWS; a political declaration underlining human control and accountability, with elements of transparency and technology review; identifying practical measures and best

69 At first glance the "sunrise slide" could be seen as excluding R&D as well as testing and evaluation from the purview of international regulation. Apart from the fact that the extent of international regulation is still to be determined – and in that sense the three regulatory regimes of industry standards, national regulations and international norms are "sliding doors" – this visualization does not exclude the penetration of international norms into domestic regimes on R&D, testing and evaluation, just as IHL weapons reviews have been internalized in domestic practice.

70 Potential carve-outs under some definitions for certain countries have also been an issue in disarmament and arms control negotiations before, and the CCW experience with the failed negotiations on a protocol on cluster munitions in 2011 must have weighed on the minds of some delegates.

71 CCW GGE on LAWS, *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc. CCW/GGE.1/2017/CRP.3, 23 October 2018, paras 24–26.

72 *Ibid.*, para. 21.

73 CCW GGE on LAWS, above note 65, para. 16(b).

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

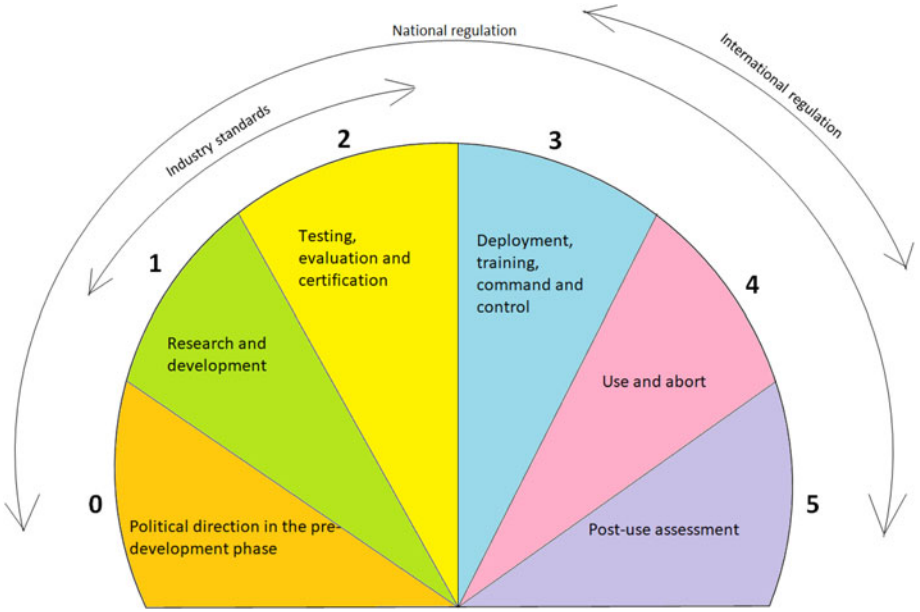

Figure 1. Touchpoints for reinforcing human involvement and oversight and for distributed governance of emerging technologies in the area of lethal autonomous weapons systems.

practices for improving compliance with international law; and a "do-nothing" approach, since IHL is fully applicable.[74]

The ten principles agreed in 2018 are applicability of IHL, non-delegation of human responsibility, accountability for use of force in accordance with international law, weapons reviews before deployment, incorporation of physical, non-proliferation and cyber security safeguards, risk assessment and mitigation during technology development, non-harm to civilian research and development (R&D) and use, the need to adopt a non-anthropomorphic perspective on AI, and the appropriateness of the CCW as a framework for dealing with the issue.[75] These are accompanied by broad understandings on the human–machine interface, on the issue of a potential definition of LAWS, and on continued review of technology pertinent to LAWS. As highlighted in Figure 1, the understandings on the human–machine interface are built around touchpoints from political direction in the pre-development phase all the way to post-use assessment. Significantly, the CCW GGE on LAWS endorsed that accountability threads together these various human–machine touchpoints in the context of the CCW. The GGE also agreed on the need to move in step with technology and

74 CCW GGE on LAWS, *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc. CCW/GGE.1/2018/3, 23 October 2018.

75 *Ibid.*, section III.A, "Possible Guiding Principles".

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

build in partnership with industry and other stakeholders a common scientific and policy vernacular across the globe.[76]

The 2019 session of the GGE added an important additional principle to the above ten, building on previous work on the human–machine interface and on the applicability of IHL:

> Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular International Humanitarian Law (IHL). In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole.[77]

It can be argued that principles might not be enough, and that they would have to be clearly linked to practice and to accountability. This criticism is valid, but given the challenges of building consensus on a complex and fast-evolving issue, a principles-first approach allows agreements to be accumulated over time without prejudice to any framework that might eventually be chosen for an international agreement.

Future directions for multilateral forums on cyber security and autonomous weapons

It is safe to make three predictions in the context of digital technologies in armed conflict. Digital technologies will continue to force formerly separate fields of technology development to merge and create new use scenarios both in the civilian and military domains.[78] It will be difficult to outlaw cyber weapons, and in fact the attack surface for malicious actors will continue to expand, for example, as use of big data for predictive personal health and precision public health expands.[79] Autonomy in technology systems will continue to grow and AI will become more of a general-purpose technology used right across the military, from recruitment and training to decision-making on force.

What directions for multilateral forums are possible in this landscape? These forums' uniqueness lies in their convening ability and their inclusiveness. Their competitive advantage is in providing a neutral platform for negotiating common principles and norms in order to clarify what is acceptable and what is

76 A. S. Gill, above note 5.

77 CCW GGE on LAWS, *Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the area of Lethal Autonomous Weapons Systems*, UN Doc. CCW/GGE.1/2019/3, 25 September 2019, section III, para. 16.

78 As the areas of genomics, emerging infectious diseases and digital technologies converge, new challenges will also emerge for multilateral forums not covered in this survey, such as the 1972 Biological and Toxin Weapons Convention. Eleonore Pauwels, "The Internet of Living Things", *Scientific American Blog*, 25 July 2017.

79 Michael Snyder and Wenyu Zhou, "Big Data and Health", *Lancet Digital Health*, Vol. 1, 29 August 2019.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

not. Today's digital challenges require these forums to expand their platform boundaries and lower entry barriers in order to bring in multiple actors, particularly technologists, industry, academia and civil society, also with a view to mitigating rising techno-nationalism. Their secretariats and office bearers need to demonstrate creativity and agility in engaging industry, technology developers and civil society. In particular, articulating a shared vision at every step is essential to drawing these different actors together.

Beyond getting the right people into the room, multilateral forums need to think about shifting their approach to the elaboration of norms, and the kind of norm-making they should prioritize in the digital age.[80] The premium placed on international treaties needs to change; instead, a flexible palette of legal, political and technical norms should be prioritized. This is not an either/or choice between general principles and binding measures, as the CCW example demonstrates. The former can lead to the latter. Further, it is not even essential that every digital issue be subject to the same set of principles. Forums can put in place their own mechanisms for discovery of foundational principles in their own context (recalling the "within the objectives and purposes" phrasing of the CCW GGE on LAWS' mandate) and then think about what approaches to implementation of these principles are feasible and effective in that context.[81] This will lead to a secondary palette of measures to "police" the norms and align action not only across nations but also across industry bodies, national governments and international bureaucracies. Again, this does not mean giving up on intrusive verification regimes wherever the stakes for compliance are very high; rather, it entails mixing and matching such measures with practical experience sharing, peer commentary and peer review to generate normative pressure.

To illustrate the above approach with the example of the CCW, the eleven principles agreed thus far in the CCW GGE on LAWS over 2017–19 can be treated as core norms.[82] These can then form the nucleus of a policy response which should include three other critical elements:

1. A set of practical understandings on national mechanisms to review potentially autonomous lethal weapons systems with regard to obligations under IHL and other applicable international law, and to rule out those systems which cannot comply with such obligations.

80 For a theoretical perspective, see Eric Talbot Jensen and Ronald T. P. Alcala (eds), *The Impact of Emerging Technologies on the Law of Armed Conflict,* Oxford University Press, New York, 2019, especially the chapter by Rebecca Crootof, "Regulating New Weapons Technology".

81 An example is the recent adoption of five principles of AI ethics by the US Department of Defense (DoD): see DoD, "DoD Adopts Ethical Principles for Artificial Intelligence", press release, 24 February 2020, available at: www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/. This should hopefully be followed by concrete measures for operationalizing the principles and clarifying their interface with legal obligations regarding AI use in weapons systems.

82 See CCW GGE on LAWS, above notes 74 and 77.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

2. A set of practical understandings on enhancing the quality of the human–machine interface in AI systems that are developed and deployed in the military domain. This would allow non-lethal system components such as decision support systems that could have a bearing on the scope and intensity of conflict to also be considered.
3. A regular technology discussion (not a review mechanism) at the CCW so that policy measures can be considered for updating in light of future technology breakthroughs. The participation of technologists and industry representatives can be formalized in this context, potentially creating a new model for multilateral forums for the participation of these stakeholders.

The implementation of the above understandings would be a national prerogative in accordance with the distributed technology governance scheme mentioned previously. Nonetheless, the voluntary experience sharing in Geneva on weapons reviews and the human–machine interface should help generate peer pressure and move all tiers of governance to higher levels of diligence and responsibility. The whole scheme can be anchored in the CCW framework creatively through a decision of the next Review Conference in 2021.

What about the cyber security forums? While it is difficult now to fundamentally re-architecture the existing forums or walk back the recent recriminations on the applicability of international law, it is possible to enlarge the scope of compromises by creating additional substantive tracks. Alongside the elaboration of core norms in the UN GGE and the OEWG, there could be parallel tracks on characteristics and use scenarios, on traceability and on cooperation and assistance. These tracks should be co-led by industry and academia representatives from different regions. On the lines of the proposed technology discussion mechanism in the LAWS context, these tracks could be a model for engaging industry and subject-matter experts on the governance of digital technologies in armed conflict.

The discussion on characteristics in the context of specific instances of cyber attacks faced by industry – in the first phase these could be focused on cyber crime or ransomware – would allow different approaches to definitions to emerge. It could also uncover common evidentiary standards, approaches to assessment of context and attenuating circumstances, and criteria for damage assessment based on financial losses or down time of critical information infrastructure.

The objective of the second track would be to address the thorny issue of attribution in a non-threatening and cooperative manner. Practically, the discussion could focus on developing an independent traceability capacity, with a roster of multidisciplinary experts who will, with the assistance of designated capacities in governments and industry, establish clear, transparent and commonly accepted standards and methodologies for collecting and analyzing evidence related to the origins and conduct of attacks, and for assessment of context and of damage resulting from the attacks.[83] When the right normative

83 While different in context, traceability will also be an important aspect of ensuring accountability in the LAWS context through national or industry-level auditability of design methodologies, training data, testing and evaluation results.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

framework is in place, traceability findings made in a neutral and apolitical setting should help facilitate peaceful resolution of disputes without finger-pointing, eliminate malware sources for which States do not claim knowledge or responsibility, plug vulnerabilities, and facilitate settlement of claims, including insurance claims for damages. Creating this capacity is a much more practical approach than creating an attribution and enforcement mechanism for which UN organizations do not have the requisite capabilities or political support.

The third track would develop methodologies for exchange of national experiences, information sharing and capacity-building through global and regional networks. While cyber crime knows no borders, protection of systems is still a strongly national exercise. Lack of trust and competing cultures and regulation further complicate information sharing among computer emergency response teams (CERTs) and between law enforcement and industry. De-politicization of the CERT function and development of technology-based trust tools or neutral third parties would facilitate information sharing on threats, coordinated disclosure of vulnerabilities and collaborative responses. When the normative framework is in place, this track could also provide an independent international assistance capacity with a roster of geographically diverse and independent experts who can provide assistance themselves or call on designated capacities in governments and industry to channel assistance to victims of massive cyber attacks.[84] This capacity should complement the efforts of networks of CERTs which act as first responders in case of attacks, and should focus on damage management and additional assistance to victims, going beyond the current mandate of national and regional CERTs.

The strategic aspects common to these three tracks are opening up the governance discussion to more actors; depoliticizing vulnerability assessment, assistance and cooperation; and de-emphasizing top-down, State-centric norm-making. In light of the schema proposed previously in this article, a tiered approach to norms is also intrinsic to these three tracks. Standards for robust software development, assessing and reporting vulnerabilities, non-riposte to cyber attacks etc. can be primarily developed at the industry level with the participation of national governments as observers. Norms for traceability, cooperation and assistance can be jointly developed for implementation through national laws, while State-centric intergovernmental forums can continue to deal with the applicability of international law and the development of new international norms. Like sliding parallel doors, these three modalities of regulation can together provide flexible responses to the challenge posed by fast-moving digital technologies to multilateral governance of digital conflict.

84 Such panels have proved to be reasonably successful in other contexts; a roster of experts has been maintained by the UN Secretary-General for chemical and biological attacks. See UN Office for Disarmament Affairs, "Secretary-General's Mechanism for Investigation of Alleged Use of Chemical and Biological Weapons", available at: www.un.org/disarmament/wmd/secretary-general-mechanism/.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press

Concluding summary

Armed conflict is taking new forms in the digital age. Both attribution and accountability are harder to establish, and the boundaries between States, the subject of regulation in multilateral forums thus far, and non-State actors are blurred. Technology itself is shifting ceaselessly and is weaponizable in unpredictable ways. Multilateral norms and norm-making practices need to adjust with agility and effectiveness. This article has highlighted a set of challenges, such as over-structured agendas, State-centrism and limited output modalities, that multilateral forums need to overcome if they are to be successful in dealing with the impact of digital technologies on international security. It has looked at recent multilateral attempts to regulate existing cyber weapons and emerging lethal autonomous weapons systems. Based on the experience of these negotiations and other trends, it has suggested a future direction for the CCW GGE on LAWS as well as for the forums dealing with cyber weapons. The key signposts for the proposed future work are interdisciplinarity, with strong engagement by the private sector; a tiered approach to norm-making, with international regulation moving in step with national regulation and industry standards; and a modular build-up of obligations anchored in guiding principles.

https://doi.org/10.1017/S1816383121000059 Published online by Cambridge University Press