2020 12th International Conference on Cyber Conflict
20/20 Vision: The Next Decade
T. Jančárková, L. Lindström,
M. Signoretti, I. Tolga, G. Visky (Eds.)
2020 © NATO CCDCOE Publications, Tallinn

R2P & Cyberspace: Sovereignty as a Responsibility

Tina J. Park, PhD
Vice President,
NATO Association of Canada
& Executive Director and Co-Founder,
Canadian Centre for the Responsibility to Protect
University of Toronto
Toronto, ON, Canada
executive.director@ccr2p.org

Michael Switzer
Deputy Executive Director
Canadian Centre for the Responsibility to Protect
University of Toronto
Toronto, ON, Canada
Michael.M.T.Switzer@gmail.com

Abstract: The Responsibility to Protect, commonly referred to as R2P or RtoP, is an emerging norm in international relations which states that when a state or government fails to protect its people from mass atrocity crimes, the international community has the responsibility to do so. First coined in 2001 and later adopted by 150 heads of state and government at the 2005 World Summit, R2P has been hailed as the most important turning point for the notion of 'sovereignty as responsibility'. Yet, to date, no proper attention has been given to understanding how the technological changes in cyberspace affect the prevention and response to R2P crimes at the national, regional and international levels. This paper explores how evolving cyber capabilities relate to the facilitation, commission and prevention of mass atrocity crimes, specifically war crimes, crimes against humanity, genocide and ethnic cleansing, under the Responsibility to Protect framework in order to (A) demonstrate that such capabilities should be examined and incorporated into the R2P discourse and (B) recommend measures to bolster the efficacy of this incorporation. It begins by discussing the historical significance of R2P, exploring its current conceptual framework and making a case for why prevention efforts deserve consideration. It then proceeds to examine three broad categories in the cyber domain (material sabotage, information collection and social influence) which may be relevant to prevention efforts of R2P. The article concludes with recommendations for more effective integration of cyber capabilities

to prevention efforts and ultimately argues that a greater attention must be given to the relationship between R2P and the cyber domain.

Keywords: *responsibility to protect (R2P), sovereignty as responsibility, prevention, mass atrocities, cyberspace*

1. INTRODUCTION

On 5 November 2018, Facebook admitted that it had failed to prevent its platform from 'being used to foment division and incite offline violence' amid the ongoing ethnic cleansing of the Rohingya people in Myanmar.[1] However, such incitement was not a random incident. It represented part of a campaign, expressed through cyber means, to create the conditions for the mass atrocity that is currently unfolding in Myanmar. As the *New York Times* reported, '[m]embers of the Myanmar military were the prime operatives behind a systematic campaign on Facebook that stretched back half a decade and that targeted the country's mostly Muslim Rohingya minority group'.[2] In fact, while the widespread use of Facebook as a platform for inciting hate may be a recent phenomenon, the use of communications technology in augmenting the commission of mass atrocity crimes is nothing new. For instance, in the build-up to the 1994 Rwandan genocide, Hutu elites used the Radio Mille Collines to incite hatred against Tutsis and Hutu moderates.[3] Once the killings began, the radio was used to relay instructions, lists of names and messages of support to *génocidaires* throughout the country.[4] In turn, the Rwandan genocide saw the most efficient and ruthless massacre of some 800,000 innocent lives over the course of merely a hundred days, while the international community remained as silent bystanders.[5]

The cases of Myanmar and Rwanda both demonstrate a well-known fact: mass atrocity crimes do not happen overnight and technology can be easily used or abused for these incidents. With proper early warning systems and efficient response mechanisms in the cyber domain, they can be prevented and halted in a timely manner. These crimes present a clear shock to values codified in the Universal Declaration of Human Rights

1 Alex Warofka, 'An Independent Assessment of the Human Rights Impact of Facebook in Myanmar,' *About Facebook*, November 5, 2018, https://about.fb.com/news/2018/11/myanmar-hria/.

2 Paul Mozur, 'A Genocide Incited on Facebook, With Posts From Myanmar's Military,' *The New York Times*, October 15, 2018, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

3 Evaina Bonnier, Jonas Poulsen, Thorsten Rogall, Miri Stryjan, 'Preparing for Genocide: Quasi-Experimental Evidence from Rwanda'. (No 31, SITE Working Paper Series from Stockholm School of Economics, Stockholm Institute of Transition Economics, 2015), 25. https://pdfs.semanticscholar.org/d78f/0bc73c715b9af13d69f9afeaedc4cbfb30bd.pdf.

4 Samantha Power, *A Problem from Hell: America and the Age of Genocide*, (New York: Basic Books, 2013), 371.

5 Power (supra n. 4), 327.

and to our collective conscience. The pledge of 'never again' has been enshrined in the Responsibility to Protect (R2P), an international norm which asserts that when sovereign states are unable or unwilling to fulfil their responsibility to protect their own populations from mass atrocity crimes, the international community has a responsibility to do so. While the military intervention aspect of R2P has been quite controversial since the inception of the principle, R2P still represents an important milestone in perceiving sovereignty as responsibility. One of R2P's primary strengths lies in its holistic approach to prevention. In promoting prevention as a key pillar of R2P, the UN Office on Genocide Prevention and the Responsibility to Protect has created a Framework of Analysis for Atrocity Crime Prevention (referred to henceforth as 'the Framework'), which helps to identify the risks for the commission of mass atrocity crimes and produces a series of indicators to guide prevention efforts.[6] In this way, R2P's operational framework focuses not only on how we may address atrocity crimes, but also on the factors that give rise to such crimes.

This article analyses the relationship between the evolution of the cyber domain and the prevention of R2P crimes – and, more importantly, how the international community may best leverage cyber capabilities to advance R2P's ultimate objective: a world free from mass atrocity crimes.

This article will underline the following key arguments: first, cyber capabilities should be incorporated into efforts to implement R2P; second, the application of R2P must be widened to include private sector partnership, especially in the prevention stage. This article is divided into three parts. First, it examines R2P's historical significance, theory and preventative utility. Second, it argues that there are already key points of intersection between cyber capabilities and R2P – which presents both challenges to and opportunities for prevention. Lastly, it argues, on the basis of this examination, for the incorporation of cyber capabilities into R2P – concluding with suggestions for moving forward.

2. WHAT IS R2P? EVOLUTION OF R2P FROM 2001 UNTIL THE PRESENT

A. Origins of R2P: The 2001 ICISS Report

To understand how R2P is different from the classic conception of humanitarian intervention, it is useful to examine the norm's origins. Following the end of the Cold War, the rise of conflicts in Bosnia and Herzegovina, Somalia, Rwanda and Kosovo gave rise to the notion of humanitarian intervention.[7] This concept proved

6 United Nations, *Early Warnings*, United Nations Office on Genocide Prevention and the Responsibility to Protect, https://www.un.org/en/genocideprevention/early-warning.shtml.

7 Jennifer Welsh, 'Authorizing humanitarian intervention,' in Richard M. Price and Mark W. Zacher eds., *The United Nations and Global Security*, (Basingstoke: Palgrave Macmillan, 2004), 177– 192.

highly controversial. When interventions were undertaken in Somalia, Bosnia and Kosovo, they were heavily criticised.[8] However, when interventions failed to take place – particularly in the case of Rwanda – such inaction came with unfathomable human costs.[9] Following debates over the unilateral NATO intervention in Kosovo in 1999, UN Secretary-General (UNSG) Kofi Annan urged UN member states to 'find a common ground' in upholding the principles of the Charter and acting in defence of our common humanity.[10] In response, the government of Canada sponsored the creation of the International Commission on Intervention and State Sovereignty (ICISS), which released its report, *The Responsibility to Protect*, in 2001.[11]

R2P, as advanced by the Commission, consisted of three key responsibilities with respect to the protection of populations: a responsibility to prevent situations in which such harm could occur; a responsibility to react to such situations; and a responsibility to rebuild following their conclusion.[12] The ICISS report marked two notable conceptual shifts. The first was a recognition that, to reconcile the debate between non-intervention and humanitarianism, it was necessary to see a state's sovereignty as implying a responsibility to protect its own population.[13] The second was a shift in the conceptual language surrounding the response to humanitarian disasters. This encompassed a change from the language of 'humanitarian intervention' which focused on the rights of intervening states, to the language of a 'responsibility to protect,' which focused on the state's duty to protect its population.[14]

B. Adoption of R2P: 2005 World Summit Outcome Document and SG's annual report on R2P

The R2P advanced by the ICISS report did not immediately take effect on the international stage, especially as the international community became occupied by the Sept 11 attacks and the 'War on Terror'. From 2001 onward, a group of 'norm entrepreneurs' came together to promote its mainstream acceptance by UN member states.[15] Their efforts met with significant success in 2005, when R2P was adopted in paragraphs 138 and 139 of the UN World Summit Outcome Document (WSOD). These paragraphs were important to the development of R2P in three respects. First,

8 International Commission on Intervention and State Sovereignty (ICISS), *The Responsibility to Protect*, (Ottawa: International Development Research Centre, 2001), Introduction.

9 Ibid.

10 United Nations Report of the Secretary General, *In Larger Freedom: Towards Development, Security and Human Rights for All*, A/59/2005, (2005), paragraph 220, https://undocs.org/A/59/2005.

11 Brian Tomlin, Norman Hillmer and Fen Hampson, *Canada's International Policies: Agendas, Alternatives and Politics*, (Toronto: Oxford University Press, 2008), 214-215.

12 International Commission on Intervention and State Sovereignty (ICISS), *The Responsibility to Protect*, (Ottawa: International Development Research Centre, 2001), xi.

13 Gareth Evans, *The Responsibility to Protect: Ending Mass Atrocity Crimes Once and For All* (Washington, DC: Brookings Institution Press, 2008), 43.

14 Ramesh Thakur, "R2P After Libya and Syria: Engaging Emerging Powers." *The Washington Quarterly* 36, no. 2 (2013), 65.

15 Tina J. Park and Victor MacDiarmid. "Selling R2P: Time For Action." In John Forrer and Conor Seyle eds., *The Role of Business in the Responsibility to Protect*, (Cambridge: Cambridge University Press, 2016), 1.

the fact that their adoption was unanimous demonstrated an international consensus on the norm. Second, great care was taken to the final articulation of R2P: the wording of paragraphs 138 and 139 were results of intense debates involving perspectives from a diversity of regions. Third, the version of R2P they advanced was different from that of the ICISS – largely due to the requirements of unanimity and compromise. While the WSOD's R2P advanced the norm's focus by constraining its scope to the four mass atrocity crimes, none of the ICISS's six criteria concerning intervention were included, nor was there any mention of a responsibility to rebuild.[16]

C. R2P Today: The Norm's Three Pillars

Since 2005, R2P has evolved as an international norm that draws on existing international law to define the responsibilities of states and the international community regarding four narrowly-defined mass atrocity crimes. R2P is an international norm in that it does not add legal obligations that constrain or determine behaviour; instead, as any norm does, it advances a shared standard of appropriate action for states, international organisations, civil society and private sector entities.[17] R2P's normative evolution is best reflected in former UN Secretary-General Ban Ki-moon's 'Three Pillar' framework. Articulated in his 2009 UN report entitled *Implementing the Responsibility to Protect*, this framework translates the commitment to R2P expressed by paragraphs 138 and 139 in the World Summit Outcome to the following three responsibilities, which are to be employed simultaneously:[18]

- **Pillar One:** Individual states have a responsibility to protect their populations from the commission and incitement of genocide, war crimes, ethnic cleansing and crimes against humanity.
- **Pillar Two:** The international community, member states, civil society and the private sector are responsible for assisting individual states in meeting their pillar one responsibilities – particularly in the context of preventing mass atrocity crimes.
- **Pillar Three:** UN member states have a responsibility to 'respond collectively in a timely and decisive manner' when a member state fails its pillar one obligations. This response must be in accordance with the 'provisions, principles and purposes' of the UN charter; while such a response could include the use of force, this measure can only be legitimate when it is authorised by the UN Security Council.[19]

These pillars form the basis of the modern conceptual understanding of R2P and are important in two aspects. First, the framework highlights a multitude of proactive

16 Yaroslav Radziwill, *Cyber-Attacks and the Exploitable Imperfections of International Law*, (Leiden: Brill Nijhoff, 2015), 288.

17 Melissa Labonte, 'R2P's Status as a Norm' in Alex J. Bellamy and Tim Dunne eds, *The Oxford Handbook of the Responsibility to Protect*, (Oxford: Oxford University Press, 2016), 137.

18 United Nations, 2009 Report of the UN Secretary-General, *Implementing the Responsibility to Protect*, A/63/677, (January 12, 2009), paragraph 12, https://undocs.org/A/63/677.

19 Ibid.

measures beyond military intervention to protect populations from atrocity crimes. The prospect of a conventional military response to the commission of atrocity crimes represents a small (albeit important) minority of the actions that R2P advocates, even in its third pillar. Alongside effective reaction, R2P prioritises a wide range of economic and diplomatic prevention methods. As such, a key strength of the R2P framework lies with the fact that it advances a set of actions that focus on ameliorating the root causes of mass atrocity crimes.

Second, the framework is 'narrow but deep' in its scope of only four, well-defined crimes.[20] Genocide, war crimes and crimes against humanity have explicit definitions in existing pieces of international law,[21] while existing *opinio juris* states that the practices that define ethnic cleansing can be assimilated into these crimes.[22] While this approach may lead to issues of contestation over applying the definition of these crimes to real-world examples, the narrowing of this scope ensures that consensus about the principle endures.[23]

D. R2P's Current Status Post-Libya: Holistic Prevention

In view of the controversial implementation of UNSC Resolution 1973 in Libya and subsequent P5 deadlock in Syria, R2P's current focus lies squarely on the strength of its holistic approach to prevention. Libya represented the first public test of R2P's implementation concerning the use of force. Before bestowing the mandate authorising NATO to use 'all necessary means to protect civilians' in resolution 1973,[24] the UN exhausted 'eleven out of the thirteen' alternative measures for which R2P advocates, including economic sanctions, preventative military deployment and arms embargoes.[25] However, as the intervention progressed, coalition leaders argued that a 'real and lasting protection of civilians could not take place with Qadhafi in power' and hence, he must be deposed.[26] This interpretation proved controversial, drawing sharp criticism from Brazil, China, India, Russia and South Africa, who charged the mission with overstepping its mandate.[27]

20 Jennifer Welsh, 'The 'Narrow but Deep Approach' to Implementing the Responsibility to Protect: Reassessing the Focus on International Crimes,' in Rosenberg, Sheri P., Tibi Galis, and Alex Zucker eds., *Reconstructing Atrocity Prevention*, (Cambridge: Cambridge University Press, 2015), 82.

21 United Nations, *Framework of Analysis*, https://www.un.org/en/genocideprevention/early-warning.shtml Annex I. Genocide is defined in Article 2 of the Convention on the Crime of Genocide; Crimes against humanity are defined in article 7 of the Rome Statute; and War Crimes are defined in article 8 of the Rome Statute.

22 United Nations, *Framework of Analysis*, 32.

23 Jennifer Welsh, 'The 'Narrow but Deep Approach' to Implementing the Responsibility to Protect: Reassessing the Focus on International Crimes,' *Op.Cit*.

24 UN Security Council Resolution 1973 (2011), S/RES/1973 (17 March 2011), https://www.undocs.org/S/RES/1973%20.

25 Paul Tang Abomo, *R2P and the US Intervention in Libya*, (New York; Palgrave Macmillan, 2018), 243.

26 Barack Obama, David Cameron and Nicolas Sarkozy, 'Libya's Pathway to Peace', *The New York Times*, 14 April 2011.

27 Alex J. Bellamy and Tim Dunne eds. *The Oxford Handbook of the Responsibility to Protect*, (Oxford: Oxford University Press, 2016), introduction.

Nevertheless, these criticisms do not translate into an outright rejection of R2P, nor do they negate the incredible degree of progress made with this emerging norm in the past few decades. Rather, the case of Libya served as a test for whether the international community could react to mass atrocity crimes in a way that solely concerned the protection of populations. As a result, little international action has been taken to stem ongoing atrocities committed by government forces in Syria. Rather, China and Russia are primarily concerned about R2P being used as a tool for regime change.[28] In this way, the 2011 intervention in Libya has merely precluded the military application of R2P's third pillar; R2P's second-pillar suite of non-military preventative measures – ranging from fostering economic stability and combating hateful ideologies to ensuring transparency in criminal justice systems – do not allow the same possibility for regime change. As such, R2P's prevention measures are far less rigid in the forms they may take, allowing for actors to find common ground, with excellent opportunities for cyber activities.

3. CYBER DOMAIN AND R2P: KEY POINTS OF RELEVANCE

To assess challenges and opportunities regarding R2P and cyber domain, this section will begin with a definition of these categories: Cyber Material Sabotage (cMS), Cyber Information Collection (cIC) and Cyber Social Influence (cSI). It will then define the ways in which each category of capability represents challenges or opportunities relevant to R2P. Cyber capabilities are defined not as technologies, but rather as the potential for an actor to effect change through a particular channel of technology in the cyber domain. This section will draw upon the UN Framework of Analysis for Mass Atrocity Prevention, to see how cyber capabilities can have implications on mass atrocity crimes.

A. Definitions: Three Categories of Cyber Capability

- **Cyber Material Sabotage (cMS)** capabilities enable an actor to damage another actor's capacity to function.
- **Cyber Information Collection & Manipulation (cICM)** capabilities enable an actor to obtain, organise and manipulate information about a population, institution, agency or operation – albeit in a way that does not cause material damage.
- **Cyber Social Influence (cSI)** capabilities enable an actor to alter the perceptions, beliefs and decision-making of a given population.

28 Thakur (supra n. 14), 71.

These three categories are not meant to exhaust the range of possibilities that an actor may realise through technological tools in the cyber domain. Instead, they are designed to allow for a more effective discussion of those sets of possibilities that are most relevant to the R2P framework.

B. Cyber Material Sabotage (cMS) Capabilities

1) Challenges

The Cyber Material Sabotage (cMS) capabilities present challenges to R2P in two ways. First is the tangible damage that cyber operations targeting financial or institutional infrastructure can cause with regards to the stability and resilience of a society. Such disruptions could constitute measures that fall under indicator 8.9 of the UN *Framework of Analysis*, namely: 'Sudden changes that affect the economy or the workforce, including as a result of financial crises, natural disasters, or epidemics'.[29] For example, in the event that any cMS capability is used to target financial services or infrastructure, it may have far-reaching consequences that could seriously disrupt the quality of life and economic stability. Examples of these disruptions are the hacking that crippled South Korean banks and infrastructure, including Korea Hydro, or the cyber-attacks on Sony and some American banks.[30]

Second, cyber thefts represent not only harm to a particular organisation or economy's capacity to function, but also an ever-growing stream of revenue that can bolster the capacities of actors to commit mass atrocities, especially collaborations with non-state terror groups. For example, since it began its cyber operations, North Korea has reportedly acquired as much as $USD 2 billion through cyber activities.[31] Much of these funds have also been successfully laundered online.[32] This revenue, in turn, has been used to develop weapons, ranging from nuclear weapons to chemical and biological weapons, which the North Korean regime sells to non-state terror groups in the Middle East such as Hezbollah and Hamas. Because of the very nature of cyber crimes and difficulties with attribution, the cMS capabilities pose real and serious threats to regional and international security, as well as day to day lives of ordinary citizens.

29 United Nations. *Framework of Analysis for Atrocity Crimes: A Tool for Prevention*. (2014). https://www.un.org/en/genocideprevention/documents/about-us/Doc.3_Framework%20of%20Analysis%20for%20Atrocity%20Crimes_EN.pdf.

30 Mattha Busby, 'North Korean 'Hacker' Charged over Cyber-Attacks against NHS,' *The Guardian* (Guardian News and Media, September 6, 2018), https://www.theguardian.com/world/2018/sep/06/us-doj-north-korea-sony-hackers-chares.

31 Edith M. Lederer, 'UN Probing 35 North Korean Cyber Attacks in 17 Countries,' *Associated Press*, (August 13, 2019), https://apnews.com/ece1c6b122224bd9ac5e4cbd0c1e1d80.

32 Michelle Nichols, 'North Korea Took $2 Billion in Cyberattacks to Fund Weapons Program: U.N. Report,' *Reuters* (August 5, 2019), https://www.reuters.com/article/us-northkorea-cyber-un/north-korea-took-2-billion-in-cyberattacks-to-fund-weapons-program-u-n-report-idUSKCN1UV1ZX.

2) Opportunities

The cMS capabilities present a limited set of opportunities for acceptable use as preventative tools. For instance, it would be illegal for states to use cMS capabilities against other states unless authorised by the UNSC, which, in the wake of Libya, may be unlikely. The *Tallinn Manual* makes it clear that: 'A State may not intervene, including by cyber means, in the internal or external affairs of another State'.[33] This encompasses (1) situations in which states intervene through cyber means and (2) situations in which states intervene in the cyber affairs of another state using non-cyber coercive means.[34] In either case, the *Tallinn Manual* asserts that 'a prohibited act of intervention' requires that 'the act in question must relate to matters that involve the internal or external affairs of the target State' and that the act 'must be coercive in nature'.[35] As cMS capabilities are, by their disruptive nature, inescapably coercive, it is unlikely that the cMS capabilities may be legitimately used by states against states.

However, there is already precedent for the use of cMS capabilities against non-state actors – for example, in 2016, the US military conducted its first offensive cyber operation against ISIS with the aim of disrupting the organisation's finances, recruiting and propaganda.[36] By reducing the financial and logistical capacity of potential non-state perpetrators, such measures take a preventative approach towards promoting the rights and freedoms of ordinary citizens. Furthermore, building a strong defence against these cMS capabilities at the national level could help foster the resilience of any society, long before any mass atrocity crimes take place.

C. Cyber Information Collection & Manipulation (cICM) Capabilities

1) Challenges

The Cyber Information Collection and Manipulation (cICM) capabilities pose challenges to R2P in two ways. First, surveillance capabilities engendered by facial recognition, GPS-tracking and the access to data transmitted through information and communication technologies (ICTs) allows actors to identify and track populations based on certain attributes, which relates to indicator 7.12 of the UN's *Framework of Analysis* by bolstering their capability to 'mark people or their property based on affiliation to a group'.[37] Indeed, the November 2019 leak of four classified Chinese bulletins illustrates the ways in which China has combined a variety of surveillance capabilities to create the Integrated Joint Operations Platform (IJOP), which included 'a detailed database of everything from an individual's exact height and electricity

33 Michael N. Schmitt et al., *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, (Cambridge: Cambridge University Press, 2017), 313.

34 Ibid.

35 Ibid. 314.

36 Ash Carter, 'A Lastine Defeat: The Campaign to Destroy ISIS,' *Harvard Belfer Center Report*, (October 2017), 32. https://www.belfercenter.org/sites/default/files/2017-10/Lasting%20Defeat%20-%20final_0.pdf.

37 United Nations, Report of the Secretary General, *Responsibility to Protect: Lessons learned for Prevention*, A/73/2019, (2019), paragraph 23, https://digitallibrary.un.org/record/3810380?ln=en.

use, to the colour of their car,' to 'if they prefer to use the front or back door to their house'. The capacity conferred by the IJOP is substantial enough that it flagged 24,412 suspicious individuals, 'of which 15,000 were sent to re-education camps and a further 706 were jailed'.[38]

Second, cyber capabilities allow for a greater degree of control over the amount and kind of information in circulation within certain sectors of cyberspace. Such cICM capabilities allow for the suppression of information relating to the early identification of mass atrocity crimes. For instance, China's 'control of court data' and 'media censorship of cases' contributed to the difficulty in assessing the extent of China's detention of journalists.[39] The capability for an actor to hide the extent of its persecution complicates prevention efforts in the way of indicator 6.11 in the UN *Framework of Analysis*, namely the 'lack of an early warning mechanism relevant to the prevention of atrocity crimes'.[40] Lastly, the emergence of 'deep fakes', such as videos generated via algorithms that make it look like a person said or did something she did not, allows actors to tamper with video evidence so as to avoid accountability, relating to indicator 3.6 of the UN *Framework of Analysis* specifically: 'Absence or inadequate external or internal mechanisms of oversight and accountability'[41] or alter a population's perception of reality through propaganda.[42]

2) Opportunities

On opportunities, the first promising cICM capability is the ability to record and monitor security forces. This capability plays a supportive role in bolstering accountability and the rule of law by providing a more transparent method of monitoring police and security forces. An example of this may be found in the creation and storage of police footage. A 2018 article by the US National Institute of Justice notes that the use of 'body-worn cameras' (BWC) by police forces may bolster transparency, allow for the storage of footage to be used as evidence in court proceedings and ensure greater capacity to refine training methods and operational strategies.[43]

All these effects may bolster the state's accountability in upholding the rule of law: knowing that police interactions are recorded may bolster the trust that the public feels towards police forces; storing police footage allows for a better capacity to hold officers accountable for their actions in a court of law; and using footage to refine

38 Emma Graham-Harrison and Juliette Garside, 'Revealed: Power and Reach of China's Surveillance Dragnet,' *The Guardian* (November 24, 2019), https://www.theguardian.com/world/2019/nov/24/china-cables-revealed-power-and-reach-of-chinas-surveillance-dragnet).

39 US Congress, Congressional-Executive Commission on China, *Annual Report 2019*, 116th Cong., 1st sess., 2019. S. Exec. Rep. 36-743, 42, https://www.cecc.gov/publications/annual-reports/2019-annual-report.

40 United Nations, *Framework of Analysis*, 15.

41 Ibid. 12.

42 Ibid. 15-16.

43 Brett Chapman, 'Body-Worn Cameras: What the Evidence Tells Us,' National Institute of Justice, Nov 14, 2018, https://nij.ojp.gov/topics/articles/body-worn-cameras-what-evidence-tells-us.

police methods may boost capacity to improve police-public relations. Indeed, even acknowledging that the BWC are present may make a positive difference. The 2016 'global, multisite randomised controlled trial'[44] study by Ariel et al. found that BWC "can reduce police use of force...when officers' discretion to turn cameras on or off is minimised".[45] Decreasing the prevalence of the use of force by police officers may boost relations between the police and the public, making it more difficult for them to be leveraged as instruments for atrocity crimes.

Second, cICM capabilities allow civilians and journalists, through smartphones and ICTs, to collect and organise media which documents risk factors for mass atrocity crimes. This is useful for effective prevention in three ways. First, ICTs can bolster prevention efforts by serving as conduits for early warning and mobilisation. For instance, during the Egyptian Revolution, protestors circumvented state censorship of the media by using smartphones to document instances of police brutality and political repression, spreading this information to international audiences and providing a wealth of evidence behind which the international community rallied.[46] Second, the ability of those undergoing active atrocity crimes to self-report enables such actors to supply a constant stream of information to policy-makers and the wider public.[47] For example, the recent leak of 24 documents relating to the Chinese internment of Muslim populations in Xinjiang has sustained broad international interest in actions that may constitute crimes against humanity.[48] Third, timely information produced by the use of digital equipment has created a new and fruitful body of potential evidence.[49] This is already the case, as both of the ICC warrants for Libyan Commander Al-Werfalli relied on videos drawn from social media.[50]

Third, while ICTs allow hate speech to propagate with ease, such speech is subject to tools of quantification and analysis. Such tools have already been applied. For example, Mondal et al. undertook a systematic measurement and analysis of hate speech on social media in 2017, allowing them to map the prevalence, targets and geographical distribution of such speech.[51] Online initiatives that have capitalised on this opportunity already exist, such as Hatebase, Islamophobic incident reporting

44 Ibid.
45 Ibid.
46 Ibid.
47 Rebecca Hamilton, 'Atrocity Prevention in the new media landscape,' AJIL Unbound, 113 (2019), 266.
48 Austin Ramzy and Chris Buckley, ''Absolutely No Mercy': Leaked Files Expose How China Organized Mass Detentions of Muslims,' *The New York Times* (The New York Times, November 16, 2019), https://www.nytimes.com/interactive/2019/11/16/world/asia/china-xinjiang-documents.html).
49 Lindsay Freeman, 'Digital Evidence and War Crimes,' *Fordham International Law Journal*, vol 41, issue 2 (2017).
50 Alexa Koenig, ''Half the Truth Is Often a Great Lie': Deep Fakes, Open Source Information and International Criminal Law,' AJIL Unbound 113 (2019): 250-255, https://doi.org/10.1017/aju.2019.47), 251.
51 Mainack Mondal et al. 'A Measurement Study of Hate Speech in Social Media,' HT '17: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, July 2017, https://homepages.dcc.ufmg.br/~fabricio/download/HT2017-hatespeech.pdf, 10.

platforms and Fight against Hate. Such capacity for observation, measurement and analysis is significant for two of the UNSG's preventative recommendations and can bolster state efforts to target hate speech. As Mondal et al. observed, their data 'might provide a unique opportunity to identify the root causes' of 'offline hate'.

Understanding the bigger picture of online hate allows for states to understand and therefore target the root causes of hatred specific to their context and thus create a more vibrant civil society, as more accurate measurement better informs action. The tools of analysis employed by Mondal et al. may also be employed to measure other indicators of respect for diversity and vibrancy of civil society. Using analytic tools to assess the field of expressed attitudes on social media may open up the possibility for governments to better understand their particular context.

D. Cyber Social Influence (cSI) Capabilities

1) Challenges

The Cyber Social Influence (cSI) capabilities represent challenges to R2P in three ways. First, they allow actors to weaken popular trust in institutions. Second, they allow actors to incite hatred and violence towards a particular group. Third, they bolster the ability for organisations committing mass atrocity crimes to recruit others to their cause, thereby enhancing their capacity to commit atrocity crimes and weaken the political will to react.

First, cSI capabilities represent a challenge to R2P by giving actors the ability to undermine the credibility of institutions, a key contributor to the outbreak of mass atrocity crimes. In that regard, influence campaigns can cast doubt on the fairness of an election, which may constitute a triggering factor under the UN *Framework of Analysis*' indicator 8.8: 'Census, elections, pivotal activities related to those processes, or measures that destabilize them'.[52] As Brangetto and Veenendaal noted, Cyber-Berkut's manipulation of voting data in Ukraine's 2014 election, while in no way influencing the election's outcome, nonetheless weakened trust in the 'credibility of the Ukrainian government in overseeing a fair election process'.[53] There is, therefore, a very important correlation between cSI capabilities and public's trust, long before any major crisis breaks out.

Second, cSI capabilities empower an actor to recruit others to their cause. This can be seen primarily in the case of online radicalisation, in which individuals are persuaded in the cyber domain to serve as an asset or an agent for a particular actor. This process has been employed by both state and non-state actors. Moreover, the

52 United Nations, *Framework of Analysis*, 17.

53 Pascal Brangetto and Matthijs Veenendaal, "Influence Cyber Operations: The Use of Cyberattacks in Support of Influence Operations", in N. Pissanidis, H. Rõigas and M. Veenendaal eds., *The 8th International Conference on Cyber Conflict*, (Tallinn: NATO CCD COE Publications, 2016), 121.

definitions of 'recruitment' and 'asset' can have different meanings depending on the situation, ranging from US citizens unwittingly cooperating with Russian Internet Research agents to organise rallies in the build-up to the US election[54] to German citizens moving to Syria to fight for ISIL.[55] Depending on the kind of recruitment, these capabilities relate to indicators such as the UN *Framework of Analysis*' indicator 7.14: 'Increased inflammatory rhetoric, propaganda campaigns or hate speech targeting protected groups, populations, or individuals' and indicator 5.3: 'Capacity to encourage or recruit large numbers of supporters from populations or groups, and availability of the means to mobilize them'.[56]

Third, cSI capabilities can be used for the incitement of hatred towards a particular group, which speaks to the UN *Framework of Analysis*' indicator 7.14: 'Increased inflammatory rhetoric, propaganda campaigns or hate speech targeting protected groups, populations or individuals' and, depending on the severity of such incitement, indicator 8.7: 'Acts of incitement or hate propaganda targeting particular groups or individuals'.[57] The effects of such influence campaigns can be seen specifically with the Myanmar military's campaign of inciting hatred against the Rohingya populations within that country's borders. As the 2018 report of the Fact Finding Commission of the Office of the High Commissioner for Human Rights concluded, hate speech 'contributed to increased tension and a climate in which individuals and groups may become more receptive to incitement'.[58] Moreover, cyber influence campaigns are highly relevant in inciting violence in general. In the case of German right-wing hate media, 'right-wing anti-refugee sentiment on Facebook predicts violent crimes against refugees in otherwise similar municipalities with higher social media usage', while violence dropped appreciably when internet access went down.[59]

2) Opportunities

These cSI capabilities present opportunities by allowing for measures which attack hate speech. Targeting hate speech includes the censorship of hateful actors through attacks on their presence in cyberspace and proactive positive engagement with their target audience. Examples of censorship include the deletion of hateful social media accounts and pages, as has been tried in the case of removing anti-Muslim

54 US House of Representatives Permanent Select Committee on Intelligence, 'Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements,' https://intelligence.house.gov/social-media-content/.

55 Pilar Cebrian, 'They Left to Join ISIS. Now Europe Is Leaving Their Citizens to Die in Iraq'. *Foreign Policy*, Sept 15, 2019. https://foreignpolicy.com/2019/09/15/they-left-to-join-isis-now-europe-is-leaving-their-citizens-to-die-in-iraq/.

56 United Nations, *Framework of Analysis*, 14.

57 Ibid., 16-17.

58 UN Human Rights Council, Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar, UN Doc. A/HRC/39/CRP.2 (Sept. 17, 2018), para. 1354.

59 Karsten Mueller and Carlo Schwarz. "Fanning the Flames of Hate: Social Media and Hate Crime," *SSRN Electronic Journal*, 2017, 1.

Facebook pages and accounts in Myanmar,[60] ISIS-affiliated twitter accounts,[61] and neo-Nazi websites.[62] These measures may be useful in short-term situations where a constrained number of actors are using online platforms to spark and coordinate violence. While censorship of media runs into a number of operational problems,[63] innovative solutions such as social media councils have been created to facilitate more proactive action against hate speech and the incitement of violence. Such councils can be useful at the early stages of a campaign of hate incitement, where populations have yet to internalise hateful messages and are therefore more open to changes in world-view. Coordination between the offices of the UN, member states and social media platforms would bolster the suppression of hateful messages and the proliferation of anti-hate campaigns. Yet, even if these measures were to be successful, they would not be effective against already-internalised hatred.

4. CONCLUSION: LOOKING FORWARD

For too long, the international community has failed in giving sufficient attention to the importance of the cyber domain in the prevention of and response to mass atrocity crimes. Yet, as this paper has demonstrated, there are many points of relevance when we consider the ways in which technology can harm or strengthen our ability to protect populations in peril. In conclusion, this article calls for the following measures.

First, a stronger partnership with the private sector, such as Google and Facebook, is a necessary first step in increasing our protection capabilities in the cyber domain. With the Artificial Intelligence (AI) revolution, it can be expected that communication within societies will increasingly rely on ICTs. In turn, companies that deliver ICT services are uniquely placed to detect and analyse warning signs, proactively remove content which incites violence and bolster international efforts to counteract the spread of hatred, whether it is through altered media, fake headlines, or inflammatory rhetoric. As such, the implementation of R2P – and, perhaps, its conceptual development – must feature buy-in from the private sector and a long-term collaboration in line of existing tenets of international law.

Second, cyber capabilities alone are not by themselves sufficient tools to prevent or halt mass atrocities; they must be combined with political leadership, existing institutions and financial, legal and social resources within a society. For example, the

60 Hannah Ellis-Petersen, 'Facebook removes accounts associated with Myanmar military,' *The Guardian*, Aug 27, 2018, https://www.theguardian.com/technology/2018/aug/27/facebook-removes-accounts-myanmar-military-un-report-genocide-rohingya.

61 'Twitter takes down 300,000 terror accounts as AI tools improve' *Financial Times*, Sept 19, 2017 https://www.ft.com/content/198b5258-9d3e-11e7-8cd4-932067fbf946.

62 https://www.telegraph.co.uk/technology/2017/08/29/worlds-oldest-neo-nazi-website-stormfront-shut/.

63 Emma Irving, 'Suppressing Atrocity Speech on Social Media,' *American Society of International Law*, (Volume 113, 2019), 260.

rapid collection of evidence – through smartphone cameras or BWC – requires that proper ethical and legal accountability measures are in place. The uses and abuses of data collected through advances in communications technology will depend on our ability to ensure that suitable protectionary measures are undertaken in the fine line between an individual's privacy and the protection of society as a whole.

Third, a long-term strategy must be devised to cope with the demands of the AI revolution in cyberspace and its impact on human rights discourse. In the near future, robots and drones could become perpetrators of crimes covered under the R2P framework, which will blur the boundaries of criminality. Unless we are able to keep up with the pace of changes brought about by the AI revolution, our pledge of 'never again' will remain a hollow promise.

Unfortunately, R2P is all-too-often dismissed as a tool for military intervention or a challenge to state sovereignty. However, at the very core of R2P is the notion of sovereignty as a responsibility. As this article has illustrated, it is important for R2P advocates and the international community at large to realise the potential that lies in proactively engaging the tools from the cyber domain. There may never be a clear blueprint for how best to prevent another genocide. Nevertheless, we all share a collective responsibility to adapt to new realities and seize new opportunities from the cyber domain.