Cyber Peace

Charting a Path Toward a Sustainable, Stable, and Secure Cyberspace

Edited by Scott J. Shackelford, Frédérick Douzet and Christopher Ankersen

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

CYBER PEACE

The international community is too often focused on responding to the latest cyber attack instead of addressing the reality of pervasive and persistent cyber conflict. From ransomware against the city government of Baltimore to state-sponsored campaigns targeting electrical grids in Ukraine and the United States, we seem to have relatively little bandwidth left over to ask what we can hope for in terms of "peace" on the Internet, and how to get there. It's also important to identify the long-term implications for such pervasive cyber insecurity across the public and private sectors, and how they can be curtailed. This edited volume analyzes the history and evolution of cyber peace and reviews recent international efforts aimed at promoting it, providing recommendations for students, practitioners, and policymakers seeking an understanding of the complexity of international law and international relations involved in cyber peace. This title is also available as Open Access on Cambridge Core.

Scott J. Shackelford is Cybersecurity Risk Management Program Chair and Executive Director of the Ostrom Workshop at Indiana University. He is also an affiliated scholar at both the Harvard Kennedy School's Belfer Center for Science and International Affairs and Stanford's Center for Internet and Society, as well as a senior fellow at the Center for Applied Cybersecurity Research.

Frédérick Douzet is Professor of Geopolitics at the University of Paris 8, Director of the French Institute of Geopolitics research team (IFG Lab), and Director of the Center for Geopolitics of the Datasphere (GEODE). She was appointed a member of the French Defense Ethics Committee in January 2020.

Christopher Ankersen is Clinical Professor of Global Affairs and Faculty Lead, Global Risk Specialization at New York University's Center for Global Affairs. He has previously worked for the United Nations and the Canadian Armed Forces.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Cyber Peace

CHARTING A PATH TOWARD A SUSTAINABLE, STABLE, AND SECURE CYBERSPACE

Edited by

SCOTT J. SHACKELFORD
Indiana University-Bloomington

FRÉDÉRICK DOUZET
University of Paris 8

CHRISTOPHER ANKERSEN
New York University

CAMBRIDGE
UNIVERSITY PRESS

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108845038
DOI: 10.1017/9781108954341

© Cambridge University Press 2022

This work is in copyright. It is subject to statutory exceptions and to the provisions of relevant licensing agreements; with the exception of the Creative Commons version the link for which is provided below, no reproduction of any part of this work may take place without the written permission of Cambridge University Press.

An online version of this work is published at doi.org/10.1017/9781108954341 under a Creative Commons Open Access license CC-BY-NC-ND 4.0 which permits re-use, distribution and reproduction in any medium for non-commercial purposes providing appropriate credit to the original work is given. You may not distribute derivative works without permission. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0

All versions of this work may contain content reproduced under license from third parties.

Permission to reproduce this third-party content must be obtained from these third-parties directly.

When citing this work, please include a reference to the DOI 10.1017/9781108954341

First published 2022

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
NAMES: Shackelford, Scott J., editor. | Douzet, Frédérick, editor. | Ankersen, Christopher, editor.
TITLE: Cyber peace : charting a path toward a sustainable, stable, and secure cyberspace / edited by Scott J. Shackelford, Frédérick Douzet, Christopher Ankersen.
DESCRIPTION: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2022. | Includes bibliographical references and index.
IDENTIFIERS: LCCN 2021058612 | ISBN 9781108845038 (hardback) | ISBN 9781108954341 (ebook)
SUBJECTS: LCSH: Cyberterrorism – Prevention. | Cyberspace – Government policy. | BISAC: LAW / International
CLASSIFICATION: LCC HV6773.15.C97 C88 2022 | DDC 363.325–dc23/eng/20220112
LC record available at https://lccn.loc.gov/2021058612

ISBN 978-1-108-84503-8 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

This volume is dedicated to our families for their ongoing support and encouragement, as well as to all those working for peace, both online and offline

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Contents

PART I BEYOND STABILITY, TOWARD CYBER PEACE: KEY CONCEPTS, VISIONS, AND MODELS OF CYBER PEACE

PART II MODALITIES: HOW MIGHT CYBER PEACE BE ACHIEVED? WHAT PRACTICES AND PROCESSES MIGHT NEED TO BE FOLLOWED IN ORDER TO MAKE IT A REALITY?

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

PART III LESSONS LEARNED AND LOOKING AHEAD

PART IV REFLECTIONS AND RESEARCH NOTES

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Contributors

Christopher Ankersen is a clinical associate professor at New York University's Center for Global Affairs, where he coordinates their Global Risk specialization. His research and teaching focus is on the fields of international security and civil–military relations. Prior to joining New York University, he worked for the United Nations in the Department of Safety and Security. His most recent publication is a co-edited volume entitled *The Future of Global Affairs: Managing Discontinuity, Disruption and Destruction.*

John K. Bonilla-Aranzales is a doctoral student in political science at the University of Missouri. His research uses a mixed methods approach to address the intersection between technology and conflict resolution mechanisms in peacebuilding scenarios. Mr. Bonilla-Aranzales is particularly interested in the Colombian case to understand how public opinion expressed in social media is related to transitional justice, truth, and reconciliation. Before starting his doctoral studies through a Fulbright Scholarship, John worked for almost five years as an advisor for strategic partnerships at the Direction of the Office of External Affairs at the University of Columbia.

Francesca Bosco She developed her expertise by focusing on cybercrime, cybersecurity, and the misuse of technology. More recently she focused on the opportunities, risks, and threats caused by new technologies. At the CyberPeace Institute she leads the development of knowledge and initiatives on disruptive technologies and how to increase resilience through capacity building.

Anne E. Boustead is an assistant professor in the School of Government and Public Policy at the University of Arizona. She researches legal and policy issues related to electronic surveillance, cybersecurity, privacy, and drug policy. She is particularly interested in empirically evaluating the impact of these policies on behavior in both the public and private sectors. She has a Ph.D. in policy analysis from the Pardee RAND Graduate School, where her dissertation was focused on the interplay between commercial data collection and law enforcement surveillance, and a JD from Fordham University School of Law.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Anne-Marie Buzatu is Vice-President and Chief Operating Officer of ICT4Peace. She is a co-founder of Security and Human Empowerment Solutions, a value-driven initiative to improve human security and development opportunities for international and national stakeholders and local communities. Prior to this, Anne-Marie was Deputy Head of the Public-Private Partnerships Division of DCAF – Geneva Centre for Security Sector Governance, Geneva, where she worked for nearly twelve years. In this role she led under a Swiss government mandate the development of the International Code of Conduct for Private Security Service Providers (ICoC), a multistakeholder initiative that set out international human rights–compliant principles and standards for the private security industry. She subsequently led the creation of the "International Code of Conduct Association" (ICoCA), the multistakeholder oversight mechanism for the ICoC, where she also served as Interim Executive Director.

Federica Carugati is a lecturer in history and political economy at King's College, London. Her research focuses on institutional development in premodern, citizen-centered governments, and on the lessons that the emergence, configuration, and breakdown of premodern institutions hold for the theory and practice of institution building today. She is the author of *A Moral Political Economy: Present, Past and Future* (Cambridge University Press, 2021) and of *Creating a Constitution: Law, Democracy, and Growth in Ancient Athens* (Princeton University Press, 2019), and her work has appeared in leading political science journals, including the *Annual Review of Political Science*, *Comparative Political Studies*, and *Perspectives on Politics*, as well as popular outlets such as *WIRED*, *The Economist*, and *la Repubblica*.

Jean-Marie Chenou is an associate professor at the Department of Political Science of the Universidad de los Andes in Bogotá (University of the Andes in Bogotá), Colombia, where he has worked since 2016. He is also a member of the board of the Red Colombiana de Relaciones Internacionales-(Colombian Network of International Relations) (Spanish REDINTERCOL) and an affiliated scholar at the Centre of International History and Political Studies of Globalization. He holds a Ph.D. in political science from the University of Lausanne in Switzerland, and an M.A. in international relations from University Paris 2 Panthéon-Assas. Before joining Los Andes, he was a lecturer at the University of Lausanne and a visiting researcher at the Department of Business and Politics at the Copenhagen Business School in Denmark. His research interests include Internet governance, the global political economy of the digital age, and the effects of digitalization on postconflict societies. His work has been published in journals such as *Colombia Internacional*, *International Journal of Transitional Justice*, *International Relations*, and *Globalizations*.

Juliana Crema has a background in political science and international relations. She holds an Erasmus Mundus Joint Master's Degree completed at Charles University, Prague; Jagiellonian University, Krakow; and Leiden University, Leiden.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

She has a range of experience across multidisciplinary areas with a primary focus on the intersection of gender, policy, and geopolitics. At the CyberPeace Institute, she is part of the advancement team, researching and analyzing how to advance the role of international law and norms in order to promote greater accountability in cyberspace.

François Delerue is a Senior Reseacher in Cybersecurity Governance at Leiden University and Project Expert on International Law for the European Cyber Diplomacy Initiative (EU Cyber Direct). He is the author of Cyber Operations and International Law (Cambridge University Press, 2020), which was awarded the 2021 Book Prize of the European Society of International Law.

Frédérick Douzet is a professor of geopolitics at the University of Paris 8, Director of the French Institute of Geopolitics research team (IFG Lab), and Director of the Geopolitics of the Datasphere (GEODE) Center. She was appointed a member of the French Defense Ethics Committee in January 2020. From 2017 to 2020, she was a commissioner of the Global Commission on the Stability of Cyberspace. In 2017, she was part of the drafting committee for the French Strategic Review of Defense and National Security. Her current research deals with the geopolitics of cyberspace, as cyberspace has become the object of power rivalries between stakeholders, a scene of confrontation, and a highly powerful tool in geopolitical conflicts. Frédérick Douzet's work aims at replacing cyber conflicts within their geopolitical context and training young researchers to take into account the cyber dimension of the geopolitical conflicts in the regions they study. She studied political science at the Institute of Political Studies of Grenoble and Oxford Brookes University. She earned a master's degree from the Graduate School of Journalism at the University of California, Berkeley in 1993, then joined the graduate school of geopolitics at the University of Paris 8 for her Ph.D. In 2015, she received the title of Chevalier de l'ordre national du Mérite in recognition of public service.

Stéphane Duguin is the CEO of the CyberPeace Institute. He has spent the last two decades analyzing how technology is weaponized against vulnerable communities. In particular, he investigated multiple instances of the use of disruptive technologies, such as artificial intelligence (AI), in the context of counterterrorism, cybercrime, cyber operations, hybrid threats, and the online use of disinformation techniques. He leads the Institute with the aim of holding malicious actors to account for the harms they cause. His mission is to coordinate a collective response to decrease the frequency, impact, and scale of cyberattacks by sophisticated actors. Prior to this position, Stéphane Duguin was a senior manager and innovation coordinator at Europol. He led key operational projects to counter both cybercrime and online terrorism, such as the European Cybercrime Centre, the Europol Innovation Lab, and the European Internet Referral Unit. He is a thought leader in digital transformation and convergence of disruptive technologies. With his work published

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

in major media, his expertise is regularly sought after by high-level panels, where he focuses on the implementation of innovative responses to counter new criminal models and large-scale abuse of cyberspace.

Tabrez Y. Ebrahim is an associate professor at California Western School of Law. He is an Ostrom visiting scholar at Indiana University; a scholar at George Mason University, Antonin Scalia Law School Center for Intellectual Property x Innovation Policy; a senior cyber law researcher at William & Mary Law School, Center for Legal & Court Technology; and a visiting fellow at the Nebraska Governance and Technology Center. He has been a visiting research fellow at Bournemouth University Centre for Intellectual Property Policy & Management in England and is a registered U.S. patent attorney. He graduated with J.D. and M.B.A. degrees from Northwestern University, an LL.M. degree from the University of Houston Law Center, an M.S. degree in mechanical engineering from Stanford University, and a B.S. degree in mechanical engineering from the University of Texas at Austin.

Camille François is the chief innovation officer at Graphika. François and her team use machine learning to map out online communities and the ways information flows through networks. They apply data science and investigative methods to these maps to find the telltale signatures of coordinated disinformation campaigns. François and colleagues at Oxford used this approach to help the US Senate Select Committee on Intelligence better understand Russian activities during and after the 2016 presidential election. She is also a Mozilla fellow, a Fulbright Scholar, and an affiliate of the Berkman Klein Center for Internet & Society.

Deborah Housen-Couriel is the Chief Legal Officer and Vice-President Regulation for Konfidas Digital Ltd., a cyber and data protection consulting firm located in Tel Aviv. Her expertise focuses on international cyber and data protection law. Her international experience includes work as a core expert on the Manual on International Law Applicable to Military Uses of Outer Space and as a Working Group Chair of the Global Forum on Cyber Expertise. Deborah was a member of the International Group of Experts that drafted the 2017 *Tallinn Manual 2.0* on state activity in cyberspace. She currently serves on the Advisory Board of the Hebrew University Law School's Cyber Security Research Center and as a research fellow with the Reichman University's Interdisciplinary Center Herzliya's Institute for Counter-Terrorism. Deborah teaches cyber law and policy at both of these universities. In 2011, she co-chaired the Regulation and Policy Committee of the National Cyber Initiative, launched by Israel's prime minister, and from 2013 to 2014 served on National Cyber Bureau's Public Committee on the Cyber Professions. She is a graduate of Harvard Kennedy School (MPA-MC), the Law School of Hebrew University (LL.B., LL.M.), and Wellesley College (*scl*).

Aude Géry holds a Ph.D. in public international law and is a postdoctoral fellow at GEODE, a research and training center on the geopolitics of the datasphere

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

hosted at the University of Paris 8. Her thesis, which was awarded the thesis prize of the French branch of the International Law Association and the third thesis prize of the IHEDN, was on international law and the proliferation of digital weapons. Her research focuses on the international regulation of digital space and more particularly on the external legal policies of States, multilateralism in the field of ICTs in the context of international security and the normative issues flowing from the adoption of instruments on digital issues. She has participated in several high-level dialogues on digital issues (Sino-European Dialogue on Cybersecurity, Track 1.5 dialogues organized by EU Cyber Direct, Global Commission on the Stability of Cyberspace, Paris Call for Trust and Security in Cyberspace) and regularly engages with state and non-state actors on these topics.

Kayle Giroud is a partnership and business development Assistant at the Global Cyber Alliance (GCA). Her role is to research and identify prospective partners in Europe and Africa, respond to their needs, and develop and manage engagement.

Benjamin Jensen's teaching and research explore the changing character of political violence and strategy. Jensen is Professor at the Marine Corps University (MCU), School of Advanced Warfighting. At MCU, he runs the advanced studies program. The program integrates student research with long-range studies on future warfighting concepts and competitive strategies in the US defense and intelligence communities. His book *Forging the Sword: U.S. Army Doctrine, 1975–2010* was published by the Stanford University Press in 2016. His second book *Cyber Strategy: The Changing Character of Cyber Power and Coercion* was published in 2018 by the Oxford University Press.

Rob Knake is a senior research scientist in Cybersecurity and Resilience at the Global Resilience Institute and the Whitney Shepardson Senior Fellow at the Council on Foreign Relations. His work focuses on Internet governance, public–private partnerships, and cyber conflict, and his expertise includes developing presidential policy. Knake served from 2011 to 2015 as Director for Cybersecurity Policy at the National Security Council. In this role, he was responsible for the development of presidential policy on cybersecurity, and built and managed federal processes for cyber incident response and vulnerability management. Knake holds a master's in public policy from Harvard's Kennedy School of Government and undergraduate degrees in history and government from Connecticut College.

Vineet Kumar is the president and founder of the Cyber Peace Foundation. He is the recipient of eight international and seventeen national awards and accolades.

Rebekah Lewis, JD, CISSP, CIPP, is a cybersecurity governance, law, and policy expert. Her diverse professional experience includes serving as a practicing attorney for the US National Security Agency and with Latham & Watkins, as a university faculty member and the director of an academic research center in Washington,

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

DC, and on cross-disciplinary teams with the World Economic Forum and the International Telecommunication Union.

Cyanne E. Loyle is an associate professor of political science and a global fellow at the Peace Research Institute Oslo. Dr. Loyle is also the co-director of the Northern Ireland Research Initiative and co-creator of the Post-Conflict Justice and During-Conflict Justice databases. She is also the co-convener of the Rebel Governance Network. Loyle's current research focuses on transitional justice adopted during and after armed conflict. Her current projects include work on rebel judicial institutions, government use and misuse of transitional justice, and digital repression. Dr. Loyle received her M.A. in holocaust and genocide studies from Stockton University and her M.A. and Ph.D. in political science from the University of Maryland.

Renée Marlin-Bennett is a professor of political science at Johns Hopkins University in Baltimore, MD, USA. Her research focuses on the nature of political power, information flows, bodies and emotions, and borders. Her publications on this theme include numerous articles in scholarly journals, such as *International Political Sociology*, *Critical Studies on Security*, *Art and International Affairs*, and *Journal of Information Technology and Politics*; and four books: *Science, Technology and Art in International Relations* (Routledge, 2019); *Alker and IR: Global Studies in an Interconnected World* (Routledge, 2012); *Knowledge Power: Intellectual Property, Information, and Privacy* (Lynne Rienner, 2004); and *Food Fights: International Regimes and the Politics of Agricultural Trade Disputes* (Gordon & Breach, 1993, republished by Routledge Revivals). From 2017 to 2019, Marlin-Bennett served as the founding editor-in-chief of the *Oxford Research Encyclopedia of International Studies*, a peer-reviewed, joint publication of the Oxford University Press and the International Studies Association. Previously, she was the general editor (2013–2016) and co-general editor (2012–2013) of the predecessor publication, *International Studies Online* (Wiley), also known as the International Studies Compendium Project. From 1987 to 2007, she was on the faculty of International Relations at the School of International Service, American University, where she served as Division Director of International Politics and Foreign Policy. She earned her doctorate in political science from the Massachusetts Institute of Technology and her B.A. *cum laude* in international relations from Pomona College.

Scott J. Shackelford serves on the faculty of Indiana University where he is the cybersecurity program chair, as well as the Executive Director of the Ostrom Workshop. He is also an Affiliated Scholar at both the Harvard Kennedy School's Belfer Center for Science and International Affairs and Stanford's Center for Internet and Society, as well as a senior fellow at the Center for Applied Cybersecurity Research. Professor Shackelford has written more than 100 articles, book chapters, essays, and op-eds for diverse publications. Similarly, Professor Shackelford's research has

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

been covered by an array of outlets, including *Politico*, *NPR*, *CNN*, *Forbes*, *Time*, the *Washington Post*, and the *LA Times*. He is also the author of *The Internet of Things: What Everyone Needs to Know* (Oxford University Press, 2020), *Governing New Frontiers in the Information Age: Toward Cyber Peace* (Cambridge University Press, 2020), and *Managing Cyber Attacks in International Law, Business, and Relations: In Search of Cyber Peace* (Cambridge University Press, 2014). Both Professor Shackelford's academic work and teaching have been recognized with numerous awards, including a Harvard University research fellowship, a Stanford University Hoover Institution national fellowship, a Notre Dame Institute for Advanced Study Distinguished Fellowship, the 2014 Indiana University Outstanding Junior Faculty Award, and the 2015 Elinor Ostrom Award.

Adam Shostack is a leading expert on threat modeling, and a consultant, entrepreneur, technologist, author, and game designer. He is an affiliate professor at the Paul G. Allen School of Computer Science & Engineering at the University of Washington, an advisor to the UK's Research Institute in Socio-Technical Security, and an advisory board member at the *Journal of Cybersecurity* and the Privacy Enhancing Technologies Symposium. He's also a member of the BlackHat Review Board, and helped create the Common Vulnerabilities and Exposure (CVE) and many other things. He currently helps many organizations improve their security via Shostack & Associates, and helps start-ups become great businesses as an advisor and mentor. While at Microsoft, he drove the Autorun fix via Windows Update, was the lead designer of the SDL Threat Modeling Tool v3, and created the "Elevation of Privilege" game. Adam is the author of *Threat Modeling: Designing for Security* and the co-author of *The New School of Information Security*.

Jessica Steinberg is an associate professor in the Department of International Studies at Indiana University. Her research focuses on the political economy of development, local politics of natural resource extraction, territorial sovereignty, and violent conflict. Her book *Mines, Communities, and States: The Local Politics of Natural Resource Extraction in Africa* (Cambridge University Press, 2019) investigates the strategic interaction between international mining firms, states, and local communities to understand different governance outcomes in regions of natural resource extraction. Her next book-length project explores the use of common-pool resources (forestry in particular) in conflict and postconflict contexts to explore the effect of common-pool resource management participation on local stability. Other areas of interest include technologies of repression, conflict events reporting, and private investment in unstable regions.

Megan Stifel is Executive Director, Americas, at the GCA and the founder of Silicon Harbor Consultants, a firm that provides strategic cybersecurity operations and policy counsel. She is a nonresident senior fellow in the Atlantic Council's Cyber Statecraft Initiative. Prior to that Megan served as Cybersecurity Policy Director

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

at Public Knowledge. Megan previously served as Director for International Cyber Policy at the National Security Council (NSC), where she was responsible for expanding the US government's information and communications technology policy abroad, involving cybersecurity, Internet governance, bilateral and multilateral engagement, and capacity building. Prior to the NSC, Ms. Stifel served in the US Department of Justice (DOJ) as Director for Cyber Policy in the National Security Division and as counsel in the Criminal Division's Computer Crime and Intellectual Property Section. Before joining DOJ, Ms. Stifel was in private practice, where she advised clients on sanctions and Foreign Corrupt Practices Act (FCPA) compliance. Before law school, Ms. Stifel worked for the US House of Representatives Permanent Select Committee on Intelligence. She received a Juris Doctorate from the Maurer School of Law at Indiana University, and a bachelor of arts in international studies and German, magna cum laude, from the University of Notre Dame.

Jennifer Trahan is a clinical professor at New York University's Center for Global Affairs where she directs the concentration in Human Rights and International Law. She has published scores of law review articles and book chapters including on the International Criminal Court's crime of aggression. Her book, *Existing Legal Limits to Security Council Veto Power in the Face of Atrocity Crimes* (Cambridge University Press, 2020) was awarded the 2020 ABILA Book of the Year Award by the American Branch of the International Law Association. She has additionally authored: *Genocide, War Crimes and Crimes Against Humanity: A Digest of the Case Law of the International Criminal Tribunal for Rwanda* (Human Rights Watch, 2010) and *Genocide, War Crimes and Crimes Against Humanity: A Topical Digest of the Case Law of the International Criminal Tribunal for the former Yugoslavia* (Human Rights Watch, 2006). She serves as one of the US representatives to the Use of Force Committee of the International Law Association and holds various positions with the American Branch of the International Law Association. She also served as an *amicus curiae* to the International Criminal Court on the appeal of the situation regarding Afghanistan, and serves on the Council of Advisers on the Application of the Rome Statute to Cyberwarfare. She was recently appointed Convenor of the Global Institute for the Prevention of Aggression.

Brandon Valeriano is a senior fellow at the Cato Institute and a distinguished senior fellow at the Marine Corps University. Dr. Valeriano has published five books and dozens of articles. His two most recent books are *Cyber War Versus Cyber Realities: Cyber Conflict in the International System* (2015) and *Cyber Strategy: The Evolving Character of Power and Coercion* (2018), both with Oxford University Press. Dr. Valeriano has written opinion and popular media pieces for such outlets as *The Washington Post*, *Slate*, *Foreign Affairs*, and *Lawfare*. He has provided testimony on armed conflict in front of both the United States Senate and the Parliament of the United Kingdom. His ongoing research explores conflict escalation, big data in

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

cyber security, the cyber behavior of revisionist actors, and repression in cyberspace. He holds a Ph.D. from Vanderbilt University.

Ryan Walsh is a graduate of the Cybersecurity Risk Management Program at Indiana University-Bloomington who interned at the Global Cyber Alliance and currently works at the U.S. Department of State.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Acknowledgments

This volume is like an iceberg: What you see reflects only a small part of the enormous efforts that lie below the surface.

The book began with a workshop co-hosted by the Ostrom Workshop of Indiana University, Geopolitics of the Datasphere (GEODE) of the University of Paris 8, and New York University's Center for Global Affairs (CGA) held in New York City. Alongside chapter authors, discussants provided their perspectives on the notion of cyber peace, including Chris Painter, Pano Yannakogeorgos, Jason Healey, Camille François, Angie Raymond, Rob Knake, and Amanda Craig Deckard. The editors would like to thank them for their insights and ideas. The editors would also like to thank CGA's Dean Vera Jelinek and its Director of Continuing Education and Public Programs Michelle D'Amico for their support.

Two additional cyber peace colloquia took place during 2020, facilitated by the Ostrom Workshop. A special thanks goes out to the team that made those events possible, including Emily Castle, Gayle Higgins, David Price, and Allison Sturgeon.

The editors would like to acknowledge the contributions made by the phalanx of graduate students from Indiana University involved in copyediting the chapter drafts, including Noah Galloway, Jalyn Rhodes, and Alexandra Sergueeva.

Finally, our thanks go out to Matt Galloway, Cameron Daddis, and the team at Cambridge University Press for their editorial and production assistance.

Funding for this volume has been graciously provided by the Ostrom Workshop, GEODE, and the Hewlett Foundation, with open access being generously supported by a gift from the Microsoft Corporation.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Introduction

Scott J. Shackelford, Frédérick Douzet,
and Christopher Ankersen[*]

In a world best described by pervasive cyber insecurity,[1] it may seem odd to discuss the prospects for cyber peace. From ransomware impacting communities around the world[2] to state-sponsored attacks on electrical infrastructure,[3] to disinformation campaigns spreading virally on social media, we seem to have relatively little bandwidth left over for asking the big questions, including: What is the best we can hope for in terms of "peace" on the Internet, and how might we get there? Yet the stakes could not be higher. McKinsey, for example, has argued that by 2022 "$9 trillion to $21 trillion of economic-value creation, worldwide, [will] depend on the robustness of the cybersecurity environment."[4]

To date, the online environment has appeared to be anything but peaceful, but there has been progress in the global drive for peace and security in cyberspace. For example, on November 12, 2018, the French President Emmanuel Macron gave a speech at the Internet Governance Forum in Paris, announcing the Paris Call for Trust and Security in Cyberspace – a multistakeholder statement of principles designed to help guide the international community toward greater cyber stability. The statement, among other things, called for action to safeguard civilian

[*] This introduction was first published in, and is adapted from, Scott J. Shackelford Inside the Drive for Cyber Peace: Unpacking Implications for Practitioners and Policymakers, UNIV. CAL. DAVIS BUS. L. J. (2021).

[1] *See, e.g.*, *The Growing Threat of Cyberattacks*, HERITAGE FOUND., www.heritage.org/cybersecurity/heritage-explains/the-growing-threat-cyberattacks (last visited Feb. 20, 2020).

[2] *See* Luke Broadwater, *Baltimore Transfers $6 Million to Pay for Ransomware Attack; City Considers Insurance Against Hacks*, BALTIMORE SUN (Aug. 28, 2019), www.baltimoresun.com/politics/bs-md-ci-ransomware-expenses-20190828-njgznd7dsfaxbbaglnvnbkgjhe-story.html; Karen Husa, *Panama-Buena Vista Union School District Computers and Phones Attacked by Ransomware*, KGET (Jan. 17, 2020), www.kget.com/news/local-news/panama-buena-vista-union-school-district-computers-and-phones-attacked-by-ransomware/.

[3] *See, e.g.*, ANDY GREENBERG, SANDWORM: A NEW ERA OF CYBERWAR AND THE HUNT FOR THE KREMLIN'S MOST DANGEROUS HACKERS 2 (2020).

[4] *See* Tucker Bailey et al., *The Rising Strategic Risks of Cyberattacks*, MCKINSEY Q. (2014), www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-rising-strategic-risks-of-cyberattacks.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

infrastructure, promote Internet access, and make democracy harder to hack.[5] On the day it was announced, more than 50 nations, "130 companies and 90 universities and nongovernmental groups," signed the Paris Call – a coalition that grew to 77 nations and over 600 companies by early 2020.[6] The goal was to leverage this widespread support to help drive interest in follow-on agreements to support "digital peace." For some, this included striving for a "Digital Geneva Convention."[7] Overall, the process was not unlike the multistakeholder journey that culminated in the 2015 Paris Climate Accord.[8] And progress has not stalled. In March 2021, for example, some 150 countries agreed, for the first time, on a draft set of cyber norms to guide state behavior in cyberspace.[9] Yet still only limited efforts have been made at even defining "cyber peace," to say nothing of how we can achieve this goal, such as by leveraging interdisciplinary social science frameworks such as polycentric governance.[10]

In an environment increasingly beset by cyber insecurity, we seek to begin laying out an agenda for how to achieve a positive cyber peace for the twenty-first century. Digital conflict and military action are increasingly intertwined, and civilian targets – private businesses and everyday Internet users alike – are vulnerable. As the Global Commission on Stability in Cyberspace makes clear, "[C]onflict between states will take new forms, and cyber-activities are likely to play a leading role in this newly volatile environment, thereby increasing the risk of undermining the peaceful use of cyberspace to facilitate the economic growth and the expansion of individual freedoms."[11]

Is the peaceful use of cyberspace possible? "Cyber peace" is difficult to define – as difficult, if not more so than its offline comparator. The term "cyber peace" seems to

[5] *See* Paris Call for Trust and Security in Cyberspace (Nov. 12, 2018), www.diplomatie.gouv.fr/IMG/pdf/paris_call_text_-_en_cle06f918.pdf.

[6] David E. Sanger, *U.S. Declines to Sign Declaration Discouraging Use of Cyberattacks*, N.Y. Times (Nov. 12, 2018), www.nytimes.com/2018/11/12/us/politics/us-cyberattacks-declaration.html; *Indiana University Among First to Endorse Paris Call for Trust and Security in Cyberspace*, IU Newsroom (Nov. 12, 2018), https://news.iu.edu/stories/2018/11/iu/releases/12-paris-call-for-trust-and-security-in-cyberspace.html; *Cybersecurity: Paris Call of 12 November 2018 for Trust and Security in Cyberspace*, France Diplomatie, www.diplomatie.gouv.fr/en/french-foreign-policy/digital-diplomacy/france-and-cyber-security/article/cybersecurity-paris-call-of-12-november-2018-for-trust-and-security-in (last visited Feb. 20, 2020).

[7] *The Need for a Digital Geneva Convention*, Microsoft (Feb. 14, 2017), https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/.

[8] See Scott J. Shackelford, *On Climate Change and Cyber Attacks: Leveraging Polycentric Governance to Mitigate Global Collective Action Problems*, 18 Vand. J. of Ent. & Tech. L. 653, 654 (2016).

[9] Josh Gold, *Unexpectedly, All UN Countries Agreed on a Cybersecurity Report. So What?*, CFR (Mar. 18, 2021), www.cfr.org/blog/unexpectedly-all-un-countries-agreed-cybersecurity-report-so-what.

[10] As originally explained by Professor Vincent Ostrom, "a polycentric political system would be composed of: (1) many autonomous units formally independent of one another, (2) choosing to act in ways that take account of others, (3) through processes of cooperation, competition, conflict, and conflict resolution." Vincent Ostrom, The Meaning of Federalism 225 (1991). The concept, though, has enjoyed wide application, including in the Internet governance context. *See* Scott J. Shackelford, Governing New Frontiers in the Information Age: Toward Cyber Peace (2020).

[11] Global Commission on the Stability of Cyberspace, https://cyberstability.org/ (last visited December 16, 2019).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

have originated during a program "at the Vatican's Pontifical Academy of Sciences in December 2008,"[12] though it was being used before that date, indeed as early as 2005 as Professor Renée Marlin-Bennett ably explores in Chapter 1. This conference, though, helped to crystallize the concept by releasing the "Erice Declaration on Principles for Cyber Stability and Cyber Peace" (Erice Declaration),[13] which called for enhanced cooperation and stability in cyberspace through promoting six principles, ranging from guaranteeing the "free flow of information" to forbidding exploitation and avoiding cyber conflict,[14] several of which mirror more recent efforts such as the 2018 Paris Call. Academic efforts at defining the term were slower still, beginning in the legal literature only in 2011. In 2011, for example, one of the first articles referencing "cyber peace" surfaced, though often only in reference to United Nations (UN) initiatives such as by the International Telecommunication Union (ITU)'s "five principles for cyber peace."[15]

From there, the term was used in the context of leveraging international law generally to improve cybersecurity, and that cyber peace should be built upon State responsibility and sovereignty, which presupposes the ability and willingness of diverse nations to detect and police cyberattacks and instability.[16] One through line from 2012 to the present, though, is the focus on protecting critical infrastructure as a key element of cyber peace.[17] Still, a core facet of the understanding throughout this time period was a negative cyber peace, e.g., managing the damage caused by cyberattacks rather than conceptualizing and planning for a more sustainable and equitable status quo.

Debate about cyber peace began to evolve by 2013. For example, the conceptual framework of polycentric governance was deployed to better contextualize the range of actors, architectures, and governance scales in play.[18] It was argued that:

[12] Jody R. Westby, *Conclusion*, *in* The Quest for Cyber Peace 112, 112 (Int'l Telecomm. Union & Permanent Monitoring Panel on Info. Sec. eds., 2011), www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-WFS.01-1-2011-PDF-E.pdf.

[13] *Id.*; *see* World Fed'n of Sci., Erice Declaration on Principles for Cyber Stability and Cyber Peace (2009), www.worldscientific.com/doi/abs/10.1142/9789814327503_0015.

[14] Henning Wegener, *A Concept of Cyber Peace*, *in* The Quest for Cyber Peace; see also *supra* note 12, at 77, 79–80.

[15] See Robert Davis, *All Our Eggs in One Cloud: The International Risk to Private Data and National Security, a Study of United States' Data Protection Law Using the International Communications Union Legislative Toolkit*, 21 Minn. J. Int'l L. Online 218, 245 (2011) (citing The ITU mission: Bringing the Benefits of ICT to all the World's Inhabitants, Int'l Telecom. Union, www.itu.int/net/about/mission.aspx [last visited Oct. 17, 2010]).

[16] For a similarly critical view of the potential role played by international law to regulate cyber operations from this period, see Michael Preciado, *If You Wish Cyber Peace, Prepare for Cyber War: The Need for the Federal Government to Protect Critical Infrastructure from Cyber Warfare*, 1 j.l. & Cyber Warfare 99, 99 (2012) (arguing that "cyber warfare cannot be policed through international treaties.").

[17] *See id.*; *In Search of Cyber Peace*: *A Response to the Cybersecurity Act of 2012*, 64 Stan. L. Rev. Online 106 (Mar. 8, 2012), www.stanfordlawreview.org/online/cyber-peace.

[18] Scott J. Shackelford, *The Meaning of Cyber Peace*, Notre Dame Inst. for Adv. Study Q. (Oct. 2013), https://ndias.nd.edu/news-publications/ndias-quarterly/the-meaning-of-cyber-peace/.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

> [C]yberpeace not as the absence of conflict, but as the creation of a network of multilevel regimes working together to promote global cybersecurity by clarifying norms for companies and countries alike to reduce the risk of conflict, crime, and espionage in cyberspace to levels comparable to other business and national security risks. Working together through polycentric partnerships, and with the leadership of engaged individuals and institutions, we can stop cyber war before it starts by laying the groundwork for a positive cyber peace that respects human rights, spreads Internet access, and strengthens governance mechanisms by fostering multi-stakeholder collaboration.[19]

As with the academy, the U.S. government has been slow to embrace the concept, in part to maintain freedom of operation in a dynamic and increasingly vital strategic environment. As the historian Jason Healey argued in 2014, "We [the U.S. government] like the fact that it is a Wild West because it lets us do more attack and exploitation."[20] The U.S. government has evolved on this matter, though the Trump administration in particular was not an aggressive promoter of multilateral engagement to promote stability in cyberspace.[21] Still, the 2020 *Cyberspace Solarium Commission Report*, which was established to "develop a comprehensive national strategy for defending American interests and values in cyberspace,"[22] did not even mention "cyber peace," though it did suggest a strategy of "layered deterrence" through eighty plus recommendations spread across six pillars that included the strengthening of norms.[23]

Despite a growing recognition of the positive role played by polycentric governance in attaining cyber peace,[24] there remains nearly as many differing conceptions of "cyber peace" as there are other related and equally amorphous terms, such as "sustainable development,"[25] or even "cyberspace" itself.[26] As Camille Francois of Harvard's Berkman Klein Center has stated, and as she expands upon in Part IV of

[19] Scott J. Shackelford, *Toward Cyberpeace: Managing Cyber Attacks through Polycentric Governance*, 62 Am. Univ. L. Rev. 1273, 1280 (2013) (cited by Bruce Schneier, Click Here to Kill Everybody 213 [2018]).

[20] Eric Chabrow, *Does U.S. Truly Want Cyber Peace?*, Bank Info Sec. (Aug. 11, 2014), www.bankinfosecurity.com/interviews/does-us-want-cyber-peace-i-2415.

[21] *See, e.g.*, Josephine Wolff, *Trump's Reckless Cybersecurity Strategy*, N.Y. Times (Oct. 2, 2018), www.nytimes.com/2018/10/02/opinion/trumps-reckless-cybersecurity-strategy.html.

[22] Chris Inglis, *The Cyberspace Solarium Commission: The International Impact*, Carnegie Endowment for Int'l Peace (Mar. 4, 2020), https://carnegieendowment.org/2020/03/04/cyberspace-solarium-commission-international-impact-event-7293.

[23] U.S. Cyberspace Solarium Commission, www.solarium.gov/ (last visited Apr. 8, 2020).

[24] *See, e.g.*, Julien Chaisse & Cristen Bauer, *Cybersecurity and the Protection of Digital Assets: Assessing the Role of International Investment Law and Arbitration*, 21 Vand. J. Ent. & Tech. L. 550, 551 (2019).

[25] The World Commission on Environment and Development: Our Common Future 37 (1987). *See also* Gabcikovo-Nagymaros Project (*Hung.* v. *Slovk.*), 1997 I.C.J. 7, 78 (Sept. 25) (defining sustainable development as "[the] need to reconcile economic development with protection of the environment").

[26] Damir Rajnovic, *Cyberspace—What Is It?*, Cisco Blog (July 26, 2012) (on file with authors).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

FIGURE 1 Cyber peace word cloud.

this edited volume, "If cyberspace is colonized by war, there is one essential question: what does cyberpeace look like?"[27]

There are many ways to answer that question, including from a positive peace perspective. Heather Roff of Johns Hopkins University, for example, has argued that "Cyber peace is the end state of cybersecurity. Yet it is not a mere absence of attacks, rather it is a more robust notion about the very conditions for security."[28] Others, such as Michael Robinson, view cyber peace through the lens of stability through stepped up active defense: "Cyber related action undertaken to preserve peace, however fragile, where fighting has been halted and to assist in implementing agreements achieved by the peacemakers."[29] Conversely, some groups see any cyberattack, however well meaning, as antithetical to the concept of cyber peace.[30] Figure 1 offers a word cloud summarizing some of the many elements embedded in the overall concept of cyber peace, pulled from influential declarations, policies, and norms.[31]

Regardless of this growing consensus on the benefits of a positive approach to cyber peace, the term escapes easy definition, which has been the case since the beginning. As the former German diplomat Henning Wegener wrote:

[27] Camille Francois, *What Is War in the Digital Realm? A Reality Check on the Meaning of "Cyberspace,"* Sci. Am. (Nov. 26, 2013), https://blogs.scientificamerican.com/guest-blog/what-is-war-in-the-digital-realm-a-reality-check-on-the-meaning-of-e2809ccyberspacee2809d/.

[28] Heather M. Roff, Cyber Peace: Cybersecurity Through the Lens of Positive Peace 3 (2016), https://static.newamerica.org/attachments/12554-cyber-peace/FOR%20PRINTING-Cyber_Peace_Roff.2fbbb0b16b69482e8b6312937607ad66.pdf.

[29] Michael Robinson et al., *An Introduction to Cyber Peacekeeping*, 114 J. Network & Comp. App. 1, 4 (2018).

[30] *See* FIfF, http://cyberpeace.fiff.de/Kampagne/DefinitionenEn (last visited Mar. 23, 2020) ("By 'cyberpeace' we understand peace in cyberspace in a very general sense: the peaceful application of cyberspace to the benefit of humanity and the environment.")

[31] These international laws and policies are discussed in Part II of Shackelford, *supra* note 1.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

> In the present context, cyber peace … is meant to be an overriding principle in establishing a 'universal order of cyberspace'. If the use of the term has more to do with politics and with political emphasis, with orienting the mind toward the right choices, then it also follows that it must remain somewhat open-ended. The definition cannot be watertight, but must be rather intuitive, and incremental in its list of ingredients.[32]

"Cyber peace," sometimes also called "digital peace,"[33] is a term that is increasingly used, but still little understood. It is clearly more than the "absence of violence" online, which was the starting point for how Professor Johan Galtung described the new field of peace studies he helped to found in 1969.[34] Similarly, Galtung argued that agreeing on universal definitions for "peace" or "violence" was unrealistic; instead, the goal should be landing on a "subjectivistic" definition agreed to by the majority.[35] In so doing, he recognized that as society and technology change, so too should our conceptions of peace and violence (an observation that's arguably equally applicable both online and offline). That is why he defined violence as "the cause of the difference between the potential and the actual, between what could have been and what is."[36]

Extrapolating from this logic, as technology advances, be it biometrics or blockchain, the opportunity cost of not acting to ameliorate suffering grows, as do the capabilities of attackers to cause harm. This highlights the fact that cyber peace is not a finish line, but rather an ongoing process of due diligence and risk management, echoing Wegener's sentiments just described. In this way, a positive cyber peace is defined here as a polycentric system that (1) respects human rights and freedoms,[37] (2) spreads Internet access along with cybersecurity best practices,[38] (3) strengthens governance mechanisms by fostering multistakeholder collaboration,[39] and (4) promotes stability and relatedly sustainable development.[40]

[32] Wegener, A *Concept of Cyber Peace*, *in* The Quest for Cyber Peace; see also *supra* note 17, at 77, 78.

[33] Microsoft, *supra*, note 7.

[34] Johan Galtung, *Violence, Peace, and Peace Research*, 6 J. Peace Res. 167, 168 (1969).

[35] *Id.*

[36] *Id.* ("[I]f a person died from tuberculosis in the eighteenth century it would be hard to conceive of this as violence since it might have been quite unavoidable, but if he dies from it today, despite all the medical resources in the world, then violence is present according to our definition.") This argument was first published, and is expanded upon, in Shackelford, *supra* note 10.

[37] See Scott J. Shackelford, *Should Cybersecurity Be a Human Right? Exploring the 'Shared Responsibility' of Cyber Peace*, 55 Stan. J. Int'l L. 155 (2019).

[38] Though, there is a case to be made that Internet access itself should be considered a human right. *See* Carl Bode, *The Case for Internet Access as a Human Right*, Vice (Nov. 13, 2019), www.vice.com/en_us/article/3kxmm5/the-case-for-internet-access-as-a-human-right.

[39] *See* Scott J. Shackelford & Amanda N. Craig, *Beyond the New 'Digital Divide': Analyzing the Evolving Role of Governments in Internet Governance and Enhancing Cybersecurity*, 50 Stan. J. Int'l L. 119 (2014).

[40] Advancing Cyberstability, Global Commission on the Stability of Cyberspace 13 (2019), https://cyberstability.org/wp-content/uploads/2020/02/GCSC-Advancing-Cyberstability.pdf ("Stability of cyberspace means everyone can be reasonably confident in their ability to use

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

These four pillars of cyber peace may be constructed by clarifying the rules of the road for companies and countries alike to help reduce the threats of cyber war, terrorism, crime, and espionage to levels comparable to other business and national security risks. This could encourage the movement along a cyber peace spectrum toward a more resilient, stable, and sustainable Internet ecosystem with systems in place to "deter hostile or malicious activity"[41] and in so doing promote both human and national security online and offline.[42] To achieve this goal, a new approach to cybersecurity is needed that seeks out best practices from the public and private sectors. This approach builds from the work of other scholars who have similarly criticized a fixation on Westphalian, national security-centric models of enhancing cybersecurity, and instead focuses on minimizing "structural forms of violence" across various governance scales and sectors.[43] Such an approach may be viewed as in keeping with the prevailing multistakeholder approach to Internet governance,[44] which is in contrast to the rise of the so-called "cyber sovereignty."[45]

A growing community of scholars, practitioners, and policymakers are looking beyond this baseline definition and are aiming at operationalizing a *positive* cyber peace, as is explored throughout this edited volume. This new drive is being supported by a growing coalition, including the governments of France and New Zealand, along with firms like Microsoft and nongovernmental organizations (NGOs) like the CyberPeace Institute, which is coming together to promote stability by leveraging codes of conduct, and emerging international standards aimed at reducing cyber insecurity and promoting cybersecurity due diligence. These stakeholders, and others, are helping to create and promote myriad related efforts, such as the Online Trust Alliance, ICT4Peace, and the CyberPeace Alliance, which are backed by major funders such as the Hewlett Foundation and the Carnegie Endowment for International Peace. The Paris Call itself is a broad statement of principles that focus on improving "cyber hygiene," along with "the security of digital products

cyberspace safely and securely, where the availability and integrity of services and information provided in and through cyberspace are generally assured, where change is managed in relative peace, and where tensions are resolved in a non-escalatory manner.")

[41] Obama White House, *The Comprehensive National Cybersecurity Initiative*, https://obamawhitehouse.archives.gov/node/233086 (last visited Nov. 10, 2017).

[42] Roff, *supra* note 29, at 3 (arguing for a human security approach to cyber peace). Yet the notion of including humans in conceptions of cyberspace and cybersecurity is nothing new. *See* James A. Winnfield, Jr., Christopher Kirchhoff, & David M. Upton, *Cybersecurity's Human Facto: Lessons from the Pentagon*, Harv. Bus. Rev. (Sept. 2015), https://hbr.org/2015/09/cybersecuritys-human-factor-lessons-from-the-pentagon, along with the work on human factors.

[43] Roff, *supra* note 29, at 3, 5.

[44] *See, e.g., Is Multistakeholderism Advancing, Dying or Evolving?* UNESCO (Jan. 6, 2018), https://en.unesco.org/news/multistakeholderism-advancing-dying-evolving; Stuart N. Brotman, *Multistakeholder Internet Governance: A Pathway Completed, the Road Ahead*, Brookings Inst. (2015), www.brookings.edu/wp-content/uploads/2016/06/multistakeholder-1.pdf.

[45] *See, e.g.,* Justin Sherman, *How Much Cyber Sovereignty Is Too Much Cyber Sovereignty?*, CFR (Oct. 30, 2019), www.cfr.org/blog/how-much-cyber-sovereignty-too-much-cyber-sovereignty.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

and services" and the "integrity of the Internet," among other topics.[46] Similarly, in the aftermath of the 2019 mass shootings at two mosques in Christchurch, New Zealand, the governments of eighteen nations – along with more than a dozen well-known technology firms such as Google and Facebook – adopted the Christchurch Call to eliminate terrorist and violent extremist content online. Yet neither of these Calls, and other related efforts, bind the participants, though they do help find common ground that could, in time, be codified into laws or other enforceable standards, and build consensus about cyber peace.

It is the goal of this edited volume to unpack this field by addressing fundamental questions including, but not limited to, what is cyber peace? What lessons can we learn from UN peacebuilding efforts, as well as the Digital Blue Helmets initiative? How does the quest for cyber peace relate to the UN's Sustainable Development Goals? What can we learn from previous historical epochs, such as the Pact of Paris? Can the drive for "cyber sovereignty" comport with cyber peace? How about leveraging national, bilateral, regional, and multilateral efforts within a polycentric framework? What lessons does the literature on regime complexes hold for promoting cyber peace?

The contributions in this edited volume feature a host of leading cybersecurity thought leaders from academia, nonprofits, and the private sector. They take a rich array of approaches, benefiting from their diverse backgrounds and experiences, at unpacking the concept of cyber peace.

OUTLINE OF THE BOOK

The book is structured as follows. It is divided into four main parts, each with several chapters. Part I is entitled "Beyond Stability, toward Cyber Peace: Key Concepts, Visions, and Models of Cyber Peace." It addresses conceptual approaches to cyber peace, extending the arguments contained in this introduction. In Chapter 1, Cyber Peace: Is That a Thing?, Renée Marlin-Bennett explores the evolution of the concepts of peace and how they might be applied in the cyber dimension. She argues that the term "positive cyber peace" remains a concept laden with contradictions and ambiguity. A number of ontological tensions challenge the understanding of and policy planning for cyber peace. Some advocates of cyber peace define it as a condition, whereas others see it as a practice or set of practices. As a condition, cyber peace is sometimes defined as a kind of peace, and at other times as something within cyberspace. Distinct modes of ontologizing cyber peace as a set of practices include cyber peace as cyber peacemaking, as maintaining the stability of information technology, and/or as cyber defense actions. As such, Marlin-Bennett argues for further attention to be paid to scholarship on the terms "cyber" and "peace," to boundary-setting distinctions between cyber peace and other social things, and to

[46] Paris Call for Trust and Security in Cyberspace, https://pariscall.international/en/.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

the implications of cyber peace metaphors. All of this, she contends, suggests areas for further honing the conceptualization of this important term.

Chapter 2, "Domestic Digital Repression and Cyber Peace," sees Jessica Steinberg, Cyanne E. Loyle, and Federica Carugati arguing that states have been quick to develop and adopt cyber capabilities that go far beyond mere surveillance and censorship. These have the potential to act as a brake on progress toward true cyber peace.

Part II is called "Modalities: How Might Cyber Peace Be Achieved? What Practices and Processes Might Need to Be Followed in Order to Make It a Reality?." It moves beyond the conceptual framework and sees chapter authors discuss what might be called their "operationalization." Deborah Housen-Couriel in Chapter 3, "Information Sharing as a Critical Best Practice for the Sustainability of Cyber Peace," aims to establish the deep dependence of cybersecurity on information sharing (IS) as a critical tool for enabling cyber peace. IS on cyber threats and their mitigation constitutes a critical best practice within many domestic regulatory regimes and is often defined as a confidence-building measure, or CBM, in key international regulatory initiatives. Moreover, Housen-Couriel reminds us of that implementation of IS as a voluntary or recommended best practice or CBM – rather than as a mandated regulatory requirement – has the dual advantage of bypassing the legal challenges of enforcement at the national level and, internationally, of achieving formal multistakeholder agreement on cyber norms. The difficulties of such normative barriers are characteristic of the contemporary cyber "lay of the land," awaiting resolution until binding cyber norms can be effectively incorporated into both domestic and international legal regimes. Housen-Couriel's chapter emphasizes that a critical condition for IS specifically, as well as for cyber peace in general, is the establishment of trust among diverse stakeholders, best undertaken through polycentric regulation.

Brandon Valeriano and Benjamin Jensen in their De-escalation Pathways and Disruptive Technology: Cyber Operations as Off-Ramps to War (Chapter 4) look at cyber military operations. They remind us that while many suggest that there are inherently revolutionary and transformational qualities of cyber operations as they relate to larger military campaigns, military revolutions are often hard to quantify and rely as much on people, processes, and institutions as they do on new capabilities. Beyond their raw military potential, emergent capabilities like cyber operations are just one among many factors that shape strategic bargaining, a process often defined more by questions of resolve and human psychology than objective power calculations about uncertain weapons. When examined empirically, one finds that cyber operations are less transformative than many believe. Cyber operations tend to augment other instruments of power and function more as shaping activities – political warfare and intelligence – than a decisive battle. Valeriano and Jensen seek to develop a theoretical logic for how strategic decision-makers factor the use of cyber operations as a tool during crisis decision-making. They assert that when posed with

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

a decision to escalate or dampen a crisis, cyber options provide decision-makers a method for signaling and low-level cost imposition that does not exacerbate tensions. Decision-makers tend to leverage cyber options as a method to manage escalation and decrease hostility. This chapter illustrates this logic through a wargame survey experiment and a case study, demonstrating the potential for cyber operations to provide an off-ramp away from war.

Jean-Marie Chenou and John K. Bonilla-Aranzales in Chapter 5, "Cyber Peace and Intrastate Conflicts: Toward Cyber Peacebuilding?," argue that intrastate armed conflict became the most frequent and deadly form of engagement in the world after the end of the Cold War. The "massification" of the use of information and communications technology (ICT) and the digitization of political activities have turned intrastate conflicts into information-centric conflicts. In this context, cyberspace can be a battlefield as well as a space to conduct peacebuilding activities. Drawing upon literatures in conflict resolution and cybersecurity, their chapter proposes a definition of cyber peacebuilding as an active concept that captures those actions that delegitimize online violence, build capacity within society to peacefully manage online communication, and reduce vulnerability to triggers that may spark online violence. Cyber peacebuilding, Chenou and Bonilla claim, can also shed light on the relationship between intrastate conflicts and global cyber peace, contributing to raise awareness about cyber threats in the Global South. The chapter uses the cases of Colombia and South Africa in order to illustrate the challenges and prospects of cyber peacebuilding organized around the four pillars of cyberspace outlined in this volume. Moreover, Chenou and Bonilla-Aranzales argue that cyber peacebuilding in the Global South is an essential element of the emergence of cyber peace as a global public good.

In Chapter 6, "Artificial Intelligence in Cyber Peace," Tabrez Ebrahim makes the case that AI is a rapidly growing technology field with significant implications for cyberspace. As such, he argues, it presents unique information technology characteristics that challenge a sustainable, stable, and secure cyber peace. AI raises new considerations for human control or lack thereof and how it may help or hinder risks. AI presents consequences for offensive and defensive cybersecurity applications and international implications in the path toward cybersingularity (Artificial General Intelligence, or AGI, that surpasses human intelligence in cybersecurity). Ebrahim contends that the use of AI in a technological cyber arms race will shape cyber peace policy. While recognizing the great deal of concern of an AI arms race leading to cybersingularity, this chapter recognizes that a complex tapestry of coordination is necessary to promote a stable information infrastructure. Focusing on the principle of shared governance, it argues that talent mobilization of global AI service corps can offset the negative impact of nation-states' economic competition to develop AGI.

Part III of the book is called "Lessons Learned and Looking Ahead" which concentrates on cases that highlight the promise and limitations of existing "real-world"

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

practices and how they could work in a cyber dimension. Jennifer Trahan, in Chapter 7 "Contributing to Cyber Peace by Maximizing the Potential for Deterrence: The Criminalization of Cyber-Attacks under the International Criminal Court's Rome Statute," examines how a cyberattack that has consequences similar to a kinetic or physical attack – causing serious loss of life or physical damage – could be encompassed within the crimes that may be prosecuted before the International Criminal Court (ICC). Trahan explains that while there is a very limited subset of cyber operations that might fall within the ambit of ICC's Rome Statute, there is value in thinking through when and how a cyberattack could constitute genocide, a crime against humanity, a war crime, or a crime of aggression. Trahan acknowledges limitations as to which attacks would be encompassed, particularly given ICC's gravity threshold, as well as the hurdle of proving attribution by admissible evidence that could meet the requirement of proof beyond a reasonable doubt. Notwithstanding such limitations, increased awareness of the largely overlooked potential of the Rome Statute to cover certain cyberattacks could potentially contribute to deterring such crimes and to reaching the goal of a state of "cyber peace."

In Chapter 8, "Trust but Verify: Diverse Verifiers Are a Prerequisite to Cyber Peace," Rob Knake and Adam Shostack claim that verification is a prerequisite for peace. Moreover, they assert: peace requires verification beyond "national technical means" or espionage. It requires mechanisms that are trusted and understood by the public. Their chapter lays out the case for a mechanism perhaps analogous to publicly operated seismographs. Seismographs detect not only earthquakes but also nuclear weapon tests. Similarly, a constellation of cyber data gathering tools, built from analogy to aviation safety programs, can provide authoritative evidence of violations and, in so doing, lead to public confidence in the state of peace.

Chapter 9, "Building Cyber Peace While Preparing for Cyber War," by Frédérick Douzet, Aude Géry, and François Delerue, serves as both a look forward and a conclusion for the volume. In it, the authors claim that since President Macron's launch of the Paris Call for Trust and Security in Cyberspace in the Fall of 2018, amidst the collapse of international cyber norm discussions in June 2017, the international community has contemplated and launched multiple initiatives to restore a multilateral dialogue on the regulation of cyberspace in the context of international security. In December 2018, two resolutions were adopted by the United Nations General Assembly (UN General Assembly) to set up the sixth Group of Governmental Experts (GGE) on the subject and a new Open-Ended Working Group (OEWG). Then, in October 2020, a Program of Action for advancing responsible state behavior in cyberspace was proposed, while two new resolutions were once again adopted by the UN General Assembly. This chapter offers an analysis of the multilateral efforts conducted over the past decade to build cyber peace in a context of proliferation of cyber conflicts and exacerbated geopolitical tensions. It studies more specifically how international law has been leveraged in UN negotiations to serve strategic objectives. Their findings show that the road to cyber peace is arduous, given the

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

will of states to preserve their ability to conduct cyber-offensive operations. In the early stages of consensus building up to 2016, traditional instruments of collective security – such as international law and non–binding norms of responsible behavior – have helped advance the discussions by providing an existing legal framework applicable to cyber operations as a basis for negotiation. However, since then, the renewed strategic competition and exacerbated geopolitical tensions have led states to engage not only in a cyber arms race but also in a competition for normative influence.

Part IV of the volume is made up of less formal, more free-flowing contributions. These chapters highlight the contributions and vision of a number of individuals and organizations to our understanding of cyber peace. Chapter 10 is an interview with Camille François, one of the pioneers of the concept of cyber peace. In it, she lays out the origin and evolution of the term in her work. In Chapter 11, Anne E. Boustead and Scott J. Shackelford explain how empirical research can do much to enhance our current understanding of cyber peace phenomena. However, they point out researchers often face significant barriers that – while not unique to cyber research – are particularly salient or difficult to overcome in this context. In this chapter, Boustead and Shackelford explore barriers commonly encountered in empirical cyber research and propose mechanisms for addressing them. When conducting empirical cyber studies, researchers may find it difficult to observe decisions made by a range of public and private actors (who may not be incentivized to publicize this decision-making), coordinate expertise across multiple domains, and systematically identify and observe members of the population of interest. In order to facilitate these processes, the authors recommend increased incentives for interdisciplinary research, public–private partnerships, and broader publication of cyber-related data.

The last three chapters in the book are written on behalf of nongovernmental organizations working in the field of cyber peace. Chapter 12, authored by Stéphane Duguin, Rebekah Lewis, Francesca Bosco, and Juliana Crema, all from the Cyber Peace Institute, note the frequent assessment that the path to cyber peace is complex, new, and ever-evolving. Although this may be true, the authors remind us, just because it poses a challenge does not mean it should not be discussed. They believe that it is time to address the question of accountability in cyberspace through the human-centric approach advocated for by cyber peace. In order for cyber peace to exist, human rights and freedoms need to be protected according to their respective contexts. Only by addressing cyber peace in this way, the authors assert, can we begin to sort through the puzzle pieces to create a framework for peace and stability in cyberspace. Chapter 13 is written by Megan Stifel, Kayle Giroud, and Ryan Walsh, all from the Global Cyber Alliance. They point out that among high-profile cybersecurity incidents over the past decade, several were reportedly the work of nation-state actors. The actors leveraged tactics, techniques, and procedures to take advantage of known vulnerabilities – technical and human – to undertake actions

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

that compromised personal information, risked human health, and paralyzed the global supply chain. Left unchecked, the scale and breadth of such actions can threaten international stability. Yet, the authors remind us that an examination of high-level cases suggests that basic cyber hygiene is an accessible and practical approach to mitigate such incidents, can enhance confidence in the use of ICT, and ultimately advance cyber peace. Vineet Kumar writes in his chapter that the Internet's potential can help people from the far corners of the earth to collaborate and share information for a common cause. However, this newfound access brings in its own set of vulnerabilities, threats, and risks. Crowdsourcing is one way to address these risks by using a systematic approach that makes use of the Internet's excellent capabilities using today's technologies. CyberPeace Corps is one such initiative, seeking collaboration from people of all backgrounds and from everywhere to maintain cyber peace by collectively combating cyber threats, cyberbullying, and cybercrime by upholding the cybersecurity triad of confidentiality, integrity, and availability of digital information resources across organizations. The final contribution comes from Anne-Marie Buzatu of ICT4Peace. She points out that Advanced Persistent Threat Groups are changing the very character of modern international conflict today, with yet to be fully appreciated consequences. While not officially acknowledged by States, these groups develop sophisticated computer algorithms – allegedly on behalf of governments – to gain unauthorized access to government or company computer systems. Here the algorithms remain undetected for extended periods, gathering information, including sensitive information, about defense capabilities and critical infrastructure control systems. The "Solarwinds" attack discovered in December 2020 vividly illustrates both the damage and the uncertainty these kinds of attacks can cause to international peace and security. Some authorities believe these cyber attacks are changing the very character of warfare, requiring changes in the thinking and approach of how to effectively defend against them. The chapter concludes by identifying some important elements to be considered in adapting international obligations and norms to the paradigm of cyber attacks.

We hope for this to be the first, and certainly not the last, volume dedicated to this important topic.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

PART I

Beyond Stability, toward Cyber Peace: Key Concepts, Visions, and Models of Cyber Peace

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

1

Cyber Peace

Is That a Thing?

Renée Marlin-Bennett

1 INTRODUCTION

This book defines "positive cyber peace" as a digital ecosystem that rests on four pillars:

> (1) respecting human rights and freedoms, (2) spreading Internet access along with cybersecurity best practices, (3) strengthening governance mechanisms by fostering multistakeholder collaboration, and (4) promoting stability and relatedly sustainable development.

These pillars merit broad support for their emphasis on justice, good governance, and diffusion of technology to bridge the so-called "digital divide." They were developed through a global vetting process over time and in different fora, and they represent views of technologists, civil society thought leaders, and representatives of intergovernmental organizations (see Permanent Monitoring Panel on Information Security of the World Federation of Scientists, 2009; Shackelford, 2014). Nevertheless, the conceptualization of cyber peace and its pillars deserves further probing. Is cyber peace really a kind of peace? International relations and global studies theories include a substantial body of literature on peace, a condition and/or a relation that is both more capacious than the pillars and, perhaps, in some ways inconsistent with them. In addition, the pillars seem to be different kinds of things. The first refers to abstractions that are instantiated in law and take form through the practices of governments. The second is a diffusion of a technology along with technical standards. The third is a preference for a certain form of governance, and the fourth once again brings up a technical issue, but then pivots to sustainability. If the pillars are supporting an edifice, they are doing so unevenly.[1] In this chapter,

[1] The critique presented in this chapter raises concerns that resonate with criticism of the concept, "global public goods," as discussed by David Long and Frances Woolley (2009). They suggest that "the concept is poorly defined, avoids analytical problems by resorting to abstraction, and masks the incoherence of its two central characteristics [the confusion of nonrivalness and nonexcludability]. The conclusion is that even if the concept of global public goods is effective rhetorically, precise definition and conceptual disaggregation are required to advance analysis of global issues."

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

I probe the ontological basis of the concept of cyber peace and uncover tensions in the meanings embedded in it.

The task begins with ontological questions about what kind of thing cyber peace is. This section draws on the definitions cyber peace advocates use to taxonomize the stated or implied assumptions about cyber peace as a condition or as a set of practices. As a condition, cyber peace is sometimes defined as a kind of peace, and at other times as something within cyberspace. Distinct modes of ontologizing cyber peace as a set of practices include cyber peace as cyber peacemaking, as maintaining the stability of information technology, and/or as cyber defense actions. The second section looks to international relations and cognate field scholarship for insight into further honing the conceptualization of cyber peace. The topics in this section include unpacking cyber as a modifier of peace, unpacking the concept of peace itself, exploring the boundaries of cyber peace by looking at how it is different from similar social things, and analyzing the implications of metaphors associated with cyber peace. The chapter concludes with a brief comment on the intent of the critique.

2 CONTENDING DEFINITIONS

The ontological question is what kind of thing is cyber peace or would it be if it were to exist?[2] Unless practitioners and scholars can come to some kind of consensus around the ontological nature of cyber peace the project risks incoherence. As cyber peace has slipped into the lexicon, beginning around 2008, the term has been used differently by the several interlocutors who draw upon it. Cyber peace is sometimes understood as a social condition or quality, sometimes as a set of practices, and sometimes as both. In this section, I interpret some core texts to tease out differences between the meanings and discuss the theoretical consequences of the differences.[3]

In drawing upon a text, I do not mean to imply that my short selections are representative of everything authors think about cyber peace, or that their definition is incorrect. Instead, I use these different articulations to show the variety of ways

[2] Thomas Hofweber (2005, p. 256) provides a pithy definition of ontology as the part of metaphysics "that tries to find out what there is: what entities make up reality, what is the stuff the world is made from?" The terms "ontology" and "ontological" in this chapter refer specifically to social ontology, the understanding of the stuff of the social world. John Searle (2006, p. 16) provides the examples of "baseball games, $20 bills, and national elections" as social things that depend on collective agreement over their ontologies. I can differentiate between professional baseball and Little League games; between $20 in US versus Canadian dollars; and among various kinds of national elections. Intersubjective agreement about the ontology of a $20 bill allows me to pay the cashier. In other words, we can agree epistemologically about how to determine whether the bills I proffer are indeed $20 bills. In Searle's formulation: "X *counts as* Y *in context* C" (2006, p. 18). But what counts as cyber peace in a given context is not a settled thing. As I argue in this chapter, inconsistent ontologies for what cyber peace is or for what it ought to encompass can work against the goal of creating a better normative framework.

[3] The insight that cyber peace is used in multiple ways is certainly not new. Wegener (2011) specifically draws out the distinctions.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

cyber peace is imagined. Highlighting the unsettledness of the essence of cyber peace is the point of the exercise.

3 THE CONDITION OF CYBER PEACE

An early use of the word "peace" in the context of cyberspace and the Internet is a 2008 forward written by the former Costa Rican president and Nobel laureate, Óscar Arias Sánchez, for the International Telecommunications Union's (ITU) report on the ITU's role in cybersecurity (Arias Sánchez, 2008). He referred to the need to promote "peace and safety in the virtual world" as "an ever more essential part of peace and safety in our everyday lives" and the urgency of creating a "global framework" to provide cybersecurity (p. 5). He implied that this safe place within cyberspace can be implemented through intergovernmental coordination around cybersecurity practices. The result would be to create the condition of feeling secure, very much along the lines of what one expects from the concept of "human security" (Paris, 2001; United Nations Development Program, 1994). Techniques, such as the adoption of cybersecurity best practices, Arias suggested, are tools that *promote* this safe world, but these tools are not themselves cyber peace. In context, it seems that peace and safety are not two separate goals but rather one: Safety *as* peace – either as a kind of peace or perhaps as a part of peace.

Ungoverned cyberspace is dangerous because of "the pitfalls and dangers of online predators" (Arias Sánchez, 2008, p. 4) who inhabit it. As a state of (albeit non-) nature, it is a Hobbesian (Hobbes, 1651) world of war and crime or, more precisely, the disposition toward violence which could break out at any time. This ungoverned, dangerous world of cyberspace is to be cordoned off and, perhaps, eliminated. Global coordination on cybersecurity is thus essential to promote the condition of safety.

Hamadoun Touré, writing in the introduction to *The Quest for Cyber Peace*, a joint publication of the ITU and the World Federation of Scientists (WFS), similarly seems to draw upon this Hobbesian view of ungoverned cyberspace when he writes that "[w]ithout mechanisms for ensuring peace, cities and communities of the world will be susceptible to attacks of an unprecedented and limitless variety. Such an attack could come without warning" (2011, p. 7). He continues, enumerating some of the devastating effects of such an attack. Touré's description suggests that conditions of cyberspace could break the security provided by the sovereign state (the leviathan) to its citizens. Violence is lurking just under the surface of our cyber interactions, waiting to break out. Touré, in a policy suggestion consistent with some liberal institutionalists' thinking in international relations, understands the potential of an international regime[4] (though he does not use that term) of agreed-upon rules that

[4] The special issue of *International Regimes*, edited by Stephen Krasner (1982), is widely viewed as the beginning of international regimes scholarship. However, Hayward Alker and William Greenberg (1977) introduced a similar concept of the same name earlier. More recent scholarship has focused on regime complexes (Alter & Raustiala, 2018).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

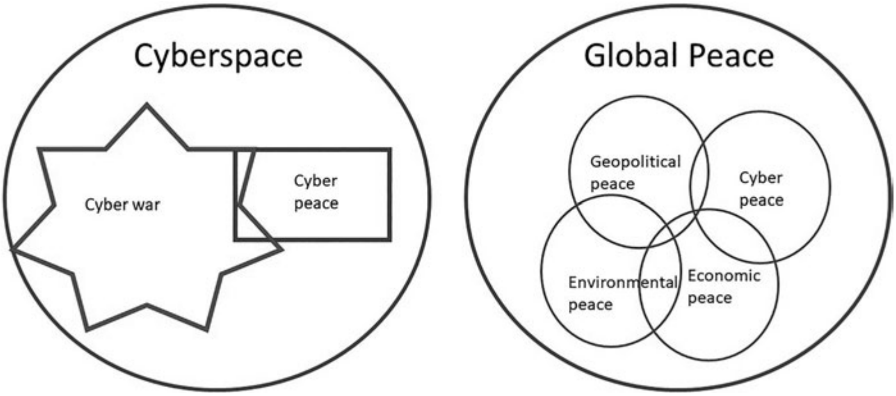


FIGURE 1.1 Different ontologies of cyber peace as conditions. On the left, both cyber peace and cyber war exist as kinds of social conditions within places of cyberspace. Cyber war is always attempting to penetrate and disrupt cyber peace. On the right, cyber peace is a subset of global peace, along with other kinds of peace.

would provide the condition of cyber peace in the absence of a single authoritative ruler. Arias and Touré both envision cyberspace as having a zone of lawlessness and war and a zone of safety and peace.

Henning Wegener's (2011) chapter in *The Quest for Cyber Peace* defines cyber peace more expansively than Touré did. More importantly, Wegener's ontology is subtly different from the division of cyberspace into the peaceful and violent zones I associated with Arias and Touré. Wegener writes:

> The starting point for any such attempted definition must be the general concept of peace as a wholesome state of tranquility, the absence of disorder or disturbance and violence – the absence not only of "direct" violence or use of force, but also of indirect constraints. Peace implies the prevalence of legal and general moral principles, possibilities and procedures for settlement of conflicts, durability and stability.
>
> We owe a comprehensive attempt to fill the concept of peace – and of a culture of peace – with meaningful content to the UN General Assembly. Its "Declaration and Programme of Action on a Culture of Peace" of October 1999 provides a catalogue of the ingredients and prerequisites of peace and charts the way to achieve and maintain it through a culture of peace (2011, p. 78).

By identifying cyber peace as a kind of peace rather than as a carve out of cyberspace, Wegener shifts the focus away from cyberspace as the world in which cyber peace exists or happens and, instead, connects to the material reality of the geopolitical world. The distinction is illustrated in Figure 1.1. The image on the left represents the definition invoked by Arias and Touré. The image on the right represents the definition invoked by Wegener.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

4 CYBER PEACE AS PRACTICES

Other interlocutors use the phrase "cyber peace" to refer to practices, which can range from using safer online platforms for cross-national communication to "cyber peace keeping" or "cyber policing" to engineering a robust, stable, and functional Internet. This approach is consistent with (though not intentionally drawing upon) what has been called the "practice turn" in international relations (Adler & Pouliot, 2011; for example, Bigo, 2011; Parker & Adler-Nissen, 2012; Pouliot & Cornut, 2015). Practices constitute meaningful social realities because of three factors. First, it matters that human beings enact practices, because in doing so we internalize that action and it becomes a part of us. Second, there is both a shared and an individual component to practices. Individuals are agentic because they can act; the action has social relevance because others act similarly. Third, practices are constituted and reconstituted through patterned behavior; in other words, through "regularity and repetition" (Cornut, 2015). Since cyber peace is an aspiration rather than something that exists now, a practice theory focus could point toward emerging or potential practices and how they are accreting.

One example of this aspirational view of practices can be found in the 2008 report, "Cyber Peace Initiative: Egypt's e-Safety Profile – 'One Step Further Towards a Safer Online Environment'," which defines cyber peace in terms of young people engaging in the practices of communicating and peacemaking.[5] According to Nevine Tewfik (2010), who summarized the findings in a presentation to the ITU, information and communications technologies (ICTs) "empower youth of any nation, through ICT, to become catalysts of change." These practices would then result in a more peaceful condition in geophysical space. Specifically, the end result would be "to create safe and better futures for themselves and others, to address the root causes of conflict, to disseminate the culture of peace, and to create international dialogues for a harmonious world" (p. 1). The report emphasized the initiative's efforts to promote safety of children online. An inference I draw from the presentation slides is that the dissemination of the culture of peace happens when children can engage safely with each other online. Cyberspace can be a place where children – perhaps because of their presumed openness to new ideas and relations – engage in peacemaking. Thus, the benefits of the prescribed cyber peace activities would spill over into the geophysical world.

Cyber peace is often defined as practices that maintain the stability of the Internet and connected services. (The tension between stability and peace will be

[5] The report on which the presentation was based is apparently no longer available online It was a joint project of Suzanne Mubarak Women's International Peace Movement, Egypt's Ministry of Communications and Information Technology, the International Telecommunication Union, and the Global Alliance for ICT and Development, in collaboration with Microsoft and Cisco Systems. The Ministry's website no longer features it, which perhaps has to do with the association of Suzanne Mubarak, or it may too old to be featured on the site. A summary of the report can be found on the website of the Virtue Foundation (Virtue Foundation Institute for Innovation and Philanthropy, n.d.).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

discussed later.) Drawing on this definition leads advocates to argue for prescriptions of protective behaviors and proscriptions of malign behaviors to maintain the functional integrity of the global ICT infrastructure. Key to this is the connection between a stable global network of ICTs and the ability to maintain peaceful practices in the geophysical world. The WFS, for example, had been concerned with all threats to information online ("from cybercrime to cyberwarfare"), but the organization's permanent monitoring panel on information security "was so alarmed by the potential of cyberwarfare to disrupt society and cause unnecessary harm and suffering, that it drafted the Erice Declaration on Principles of Cyber Stability and Cyber Peace" (Touré & Permanent Monitoring Panel on Information Security of the World Federation of Scientists, 2011, p. vii). The declaration states: "ICTs can be a means for beneficence or harm, hence also as an instrument for peace or for conflict" and advocates for "principles for achieving and maintaining cyber stability and peace" (Permanent Monitoring Panel on Information Security of the World Federation of Scientists, 2009, p. 111). These principles about how to use ICTs are, in fact, practices. By adhering to the principles and acting properly, engagements in cyberspace and ICTs promote peace in the world. The declaration seems to refer to a general condition combining life as normal without the disruptions that warlike activities cause to "national and economic security," and life with rights, that is human and civil rights, "guaranteed under international law."

In other words, for this declaration stability is a desired characteristic of cyberspace and peace is a desired characteristic of life in the world as a whole. However, it does not follow that stability is inherently peaceful, unless peace is tautologically defined as stability. The absence of cyber stability might harm peace and the presence of cyber stability might support peace, but the presence of stability is not itself peaceful, nor does it generate peace.[6] At best, we can say that peace is usually easier to attain under conditions of stability.

Another text focusing on cyber peace as a set of practices is the Cyberpeace Institute's website. It first calls for "A Cyberspace at Peace for Everyone, Everywhere," which seems to hint at cyber peace as a condition of global society, but the mission of the organization is defined primarily as the capacity to respond to attacks, and only secondarily as strengthening international law and the norms regarding conflictual behavior in cyberspace. Indeed, defense capacity is emphasized in the explanation that "The CyberPeace Institute will focus specifically on enhancing the stability of cyberspace by supporting the protection of civilian infrastructures from sophisticated, systemic attacks" (CyberPeace Institute – About Us, 2020). The ability to mount a swift defense in response to an attack does not create peace, it simply means that our defenses may be strong enough

[6] The use of cyber weapons by human rights activists to counter oppressive regimes is discussed in the section on boundaries (Section 7).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

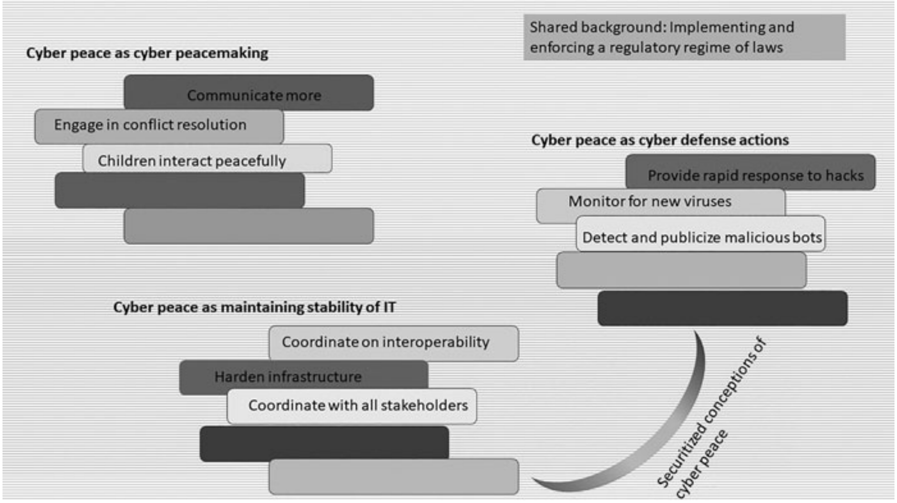

Shared background: Implementing and enforcing a regulatory regime of laws
Cyber peace as cyber peacemaking
Communicate more
Engage in conflict resolution
Children interact peacefully
Cyber peace as cyber defense actions
Provide rapid response to hacks
Monitor for new viruses
Detect and publicize malicious bots
Cyber peace as maintaining stability of IT
Coordinate on interoperability
Harden infrastructure
Coordinate with all stakeholders
Securitized conceptions of cyber peace

FIGURE 1.2 Cyber peace as the sum of practices in both securitized and non-securitized conceptualizations, against a shared background of implementing and enforcing a regulatory regime of laws.

that the attacks do not disrupt the stability of the Internet and other information technologies.

These conceptualizations of cyber peace as collections of practices thus ontologize kinds of cyber peace, which are distinct, but comparable. By comparing them, we can see underlying tensions regarding what can be considered peaceful – Is it peace making or securitization (defense and stability)? – though, as noted in the descriptions above, no collection of practices is wholly of one type. Figure 1.2 depicts different collections of practices that have been bundled together as the definition of cyber peace. (For clarity, I have not shown overlaps.) All of these conceptualizations are proposed against a background of a regulatory regime of implementing and enforcing laws.

5 CYBER PEACE AS BOTH CONDITIONS AND PRACTICES

A third category blends conditions and practices, seeing the condition of cyber peace emerge as greater than the sum of its constituent parts, which are practices. In an early iteration of his work on this concept, Scott Shackelford (2014) paints this sort of hybrid picture of cyber peace. He claims that the practices of polycentric governance related to cybersecurity spill over into a positive cyber peace:

> Cyber peace is more than simply the inverse of cyber war; what might a more nuanced view of cyber peace resemble? First, stakeholders must recognize that a positive cyber peace requires not only addressing the causes and conduct of cyber

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

> war, but also cybercrime, terrorism, espionage, and the increasing number of incidents that overlap these categories (p. 357).

This can happen, Shackelford suggests, through a process of building up governance on limited problems, thereby proliferating the number of good governance practices. The polycentric governance model specifically rejects a top-down monocentric approach:

> [A] top-down, monocentric approach focused on a single treaty regime or institution could crowd out innovative bottom-up best practices developed organically from diverse ethical and legal cultures. Instead, a polycentric approach is required that recognizes the dynamic, interconnected nature of cyberspace, the degree of national and private-sector control of this plastic environment, and a recognition of the benefits of multi-level action. Local self-organization, however – even by groups that enjoy legitimacy – can be insufficient to ensure the implementation of best practices. There is thus also an important role for regulators, who should use a mixture of laws, norms, markets, and code bound together within a polycentric framework operating at multiple levels to enhance cybersecurity (p. 359, notes omitted).

These interconnected, overlapping, small to medium-scale governance practices build upward in Shackelford's model and could eventually become a thick cybersecurity regime. When the regime is thick enough, cyber peace obtains. This model relies on a securitized notion of cyber peace, despite the discussion in the text of positive cyber peace that is more far-reaching than just the absence of war. His more recent work, co-authored by Amanda Craig, expands cyber peace to include global peace-related issues and practices, including development and distributive justice. They write:

> Ultimately, "cyber peace" will require nations not only to take responsibility for the security of their own networks, but also to collaborate in assisting developing states and building robust regimes to promote the public service of global cybersecurity. In other words, we must build a positive vision of cyber peace that respects human rights, spreads Internet access alongside best practices, and strengthens governance mechanisms by fostering global multi-stakeholder collaboration, thus forestalling concerns over Internet balkanization (Shackelford & Craig, 2014, p. 178, note omitted).

Figure 1.3 depicts this model of best practices developed from the ground up, ultimately producing a kind of cyber peace that exceeds the summation of all the different practices.

The point of this exercise of categorizing different definitions of cyber peace is to say that a definitional consensus has not been reached and to remind ourselves that the ontology built into our definitions matters for how we think about what sounds like a very good goal. Moreover, ontological foundations matter for how the practitioners among us craft policies in pursuit of that goal.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

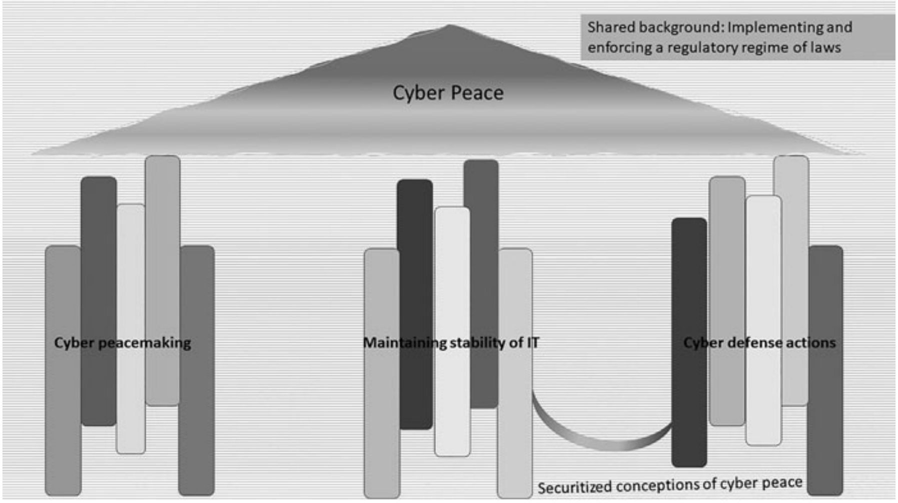

Shared background: Implementing and enforcing a regulatory regime of laws
Cyber Peace
Cyber peacemaking
Maintaining stability of IT
Cyber defense actions
Securitized conceptions of cyber peace

FIGURE 1.3 Peaceful practices and shared background emerge as cyber peace, which is greater than the sum of its parts.

6 HONING THE CONCEPT OF CYBER PEACE

The four parts of this section critically engage further with cyber peace, pointing to conceptual elements that could be productively honed to make a sharper point. The point here is not to provide an answer of what cyber peace is or should be but, rather, to draw upon scholarship from international relations and cognate fields to uncover contradictions and missed implications of the current usage. I begin by taking a closer look at "cyber" and "peace" and then turn to the boundaries of cyber peace as a social thing, followed by a discussion of the consequences of some of the metaphors associated with cyber peace.

6.1 *Unpacking the "Cyber" Element*

"Cyber" is a shortening of "cybernetics," a term introduced by Norbert Wiener, who used it to refer to the control of information machines and human groups. He emphasized: "*Cybernetics takes the view that the structure of the machine or of the organism is an index of the performance*" (Wiener, 1988, p. 57; italics in original) because the structures – that is, the properties of the machine or organism – determine what the machine or organism is able and unable to do, and what it is permitted to do, must do, and must not do. Cybernetics concerns control and order; its purpose is to be a bulwark against disorder and entropy. The shortened form quickly came to connote that which involves computers and information technology. "Cyberspace," famously introduced in *Neuromancer* by William Gibson (1994), rapidly became the narrative means of reimagining a communications technology (the Internet) as a place

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

(albeit a heterotopia [Foucault, 1986; Piñuelas, 2008]) *in* which or *on* which people (reimagined as users) do things and *to* which they go. As discussed in the section on cyberspace as a condition, we then imagine cyberspace to be a state of (non-) nature apart from the real-life physical world we live in, and we think of it as dangerous because it is ungoverned or incompletely governed. Some instances of cybercrime give credence to that, though such crimes may well be subject to law enforcement by real-life police or others. The irony is that although the cyber refers to the *realization of control*, cyberspace is thought of as a place of *lack of control*, as David Lyon (2015) has recognized.

More recent morphing of the usage of "cyber" turns it into a noun associated with military activity using information technology-intensive tools. This particular nominalization immediately calls to mind warnings from securitization theory (Balzacq, 2005; inter alia, Buzan, 1993; Hansen, 2000; Waever, 1996). The theory focuses on how language constrains our thinking and specifically on how language recasts situations, people, processes, relations, etc. as security threats, and leads to a creeping expansion of control by institutions that command the use of force. This should be understood as a danger rather than a deterministic outcome,[7] and I am not arguing that we should excise "cyber" from the dictionary. But I am mindful of the securitizing language that drags the concept of cyber peace back toward a sort of negative peace. As Roxanna Sjöstedt (2017) puts it, "If you construct a threat image, you more or less have to handle this threat."

In short, "cyber" is complicated. The word connotes the constitution of a space outside our ordinary existence in geographical space. Cyber implies order in the form of efficient control through code and other engineered rules that ought to work well. Yet cyber also hints at disorder and even chaos, since rules are often circumvented. Additionally, the military's appropriation of cyber as a shortening of "cyber conflict" or "cyber war" risks turning cyber peace into an oxymoron, taking on the sense of martial peace. That linguistic change may condition thinking and securitize the very thing that ought to be desecuritized.

6.2 *Unpacking the "Peace" Element*

If anything, peace is even more complicated than cyber. Peace is the main focus of the entire field of peace studies, and it is also an important topic for scholars of conflict management and conflict resolution, as well as of international relations more

[7] Jan Ruzicka (2019) points out the problem of case selection bias in empirical studies of securitization, with studies of successful instances of securitization being more common in the scholarship than studies of failures. A notable exception is Myriam Dunn Cavelty's (2013) nuanced study of cyber insecurity and the multiple ways it is framed. Although she leaves open the possibility of further non-securitizing responses to threats of cyberspace, she concludes that "the stronger the link between cyberspace and a threat of strategic dimensions becomes, the more natural it seems that the keeper of the peace in cyberspace should be the military" (p. 119).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

broadly. Peace always sounds good – better than war, at any rate.[8] But the war-peace dichotomy may hide the definitional complexity. Johann Galtung differentiates between "negative peace," understood as the absence of violence in a relationship and "positive peace," a more complex term that is often used to refer to relations that are just, sustainable, and conducive human flourishing in multiple ways (see also Shackelford, 2016). In its most expansive connotation, the relationship of positive peace is tied to peacebuilding and, ultimately, to amity. The main thrust of this volume envisions cyber peace as positive cyber peace. But the caveats articulated by Paul Diehl (2016, 2019) about positive peace and its usefulness as a social type of thing are worth considering. He notes, first of all, the lack of consensus among positive peace researchers about what is actually included in it:

> Conceptions include, among others, human rights, justice, judicial independence, and communication components. Best developed are notions of "quality peace," which incorporate the absence of violence, but also require things such as gender equality in order for societies to qualify as peaceful (2019).
>
> The lack of clarity over what positive peace is has, Diehl suggests, epistemological consequences.
>
> Many of [the things that are required for societies to qualify as peaceful], however, lack associated data and operational indicators. Research on positive peace is also comparatively underdeveloped (2019).[9]

While Diehl finds the concept of positive peace desirable, he warns that the concept is underdeveloped in three important ways, and each of these resonates with considerations about cyber peace.

First, what are the dimensions of peace and why is so little known about how the many dimensions interact? His concern should provoke cyber peace theorists to consider whether the four pillars are dimensions in Diehl's terms and, if so, whether they comprise *all* the dimensions. Given the potential for multiple dimensions of peace, perhaps only some are required for the situation to be deemed peaceful.

[8] Though I called out the martial quality of "cyber" (discussed in the section on cyber), one could argue that "peace" is as likely to make "cyber" seem *less* military as "cyber" is likely to make "peace" sound *more* military.

[9] To be fair, Diehl is interested in identifying better ways of understanding and studying positive peace and not just critiquing the deficiencies. It is also important to note that Diehl takes a mainstream (neopositivist) approach to social science methodology. He is concerned about operationalizing positive peace in ways that will allow researchers to subject it to mainstream hypothesis testing. That is, he is less interested in critical interpretivist approaches adopted by many theorists outside the mainstream. (Theorists outside the mainstream include those working on critical, feminist, and green theories, to name a few.) Many of the scholars writing about positive peace ally with the non-mainstream camp (to borrow a rather warlike metaphor). Disagreements over appropriate methodological approaches to research notwithstanding, most scholars will likely agree that conceptual clarity is necessary for good research, and that is the key point that Diehl is making. (Herbert Reid and Ernest Yanarella [1976] offhandedly made just that point for research on positive peace, p. 340, n. 107.) I would add that conceptual clarity is similarly necessary for advocacy based on that concept.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Alternatively, perhaps cyber peace is actually an ideal type, and the different dimensions make a situation more or less cyber peaceful.

Second, Diehl also raises the concern about an undertheorized assessment of how positive peace varies across all forms of social aggregation ("levels of analysis" in international relations scholarship). How does positive peace manifest differently in different contexts? For cyber peace, this critique points to the not fully developed idea of how the scale works in cyberspace and how that matters. A neighborhood listserv is different from Twitter, but shares some characteristics relevant to peace – flame wars and incivility are a problem in both environments. But the risks of manipulation of communication by foreign adversaries on Twitter and the kinds of policies that would be required to make peace on Twitter means, I suggest, that the environment of cyberspace is similarly complicated with regard to scale.

Third, Diehl (2019) notes that "some positive peace concepts muddle the distinction between the definitional aspects of peace and the causal conditions needed to produce peaceful outcomes." I think that the four cyber peace pillars may fall prey to this lack of conceptual clarity and, perhaps, to a sort of tautology.

7 BOUNDARIES

The next topic is boundaries and the distinctions that create them. An argument can be made that we are witnessing the creation of cyber peace as a new social entity, a thing. Andrew Abbott (1995) suggests new things emerge through a process of yoking together a series of distinctions. This is an iterative process of asking what are the characteristics of the new thing and what are not? "Boundaries come first, then entities" (p. 860). Cyber peace has yet to cohere into the sort of enduring, reproducing institution that would count as one of the Abbott's new social entities, but we do see the setting of "proto-boundaries" that may become stable when we examine the processes of trying to name and implement cyber peace. In this section, I discuss three "points of difference" that are important for the concretization of cyber peace: Between (1) cyber peace and cyber aggression, (2) cyber peace/aggression and cyber lawfulness/crime, and (3) associating multistakeholder cyber governance with cyber peace and (implicitly) associating other forms of cyber governance with non-cyber peace.

A basic distinction is between the common sense understanding of what constitutes cyber peace versus cyber aggression. The case of the 2007 cyberattack against Estonia is a clear example of cyber aggression. A more complicated case is Stuxnet, the malicious computer worm discovered in 2010, which was deployed against computer equipment used in the Iranian nuclear program. One interpretation of the Stuxnet operation would name it cyber aggression. A different interpretation would find the use of this cyber weapon de-escalatory when considered in its broader geopolitical context. Stuxnet decreased the rapid ramping up of Iran's ability to develop nuclear arms, which made an attack with full military force unnecessary. On the

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

one hand, information technology was used for a hostile purpose. On the other, the targeted cyber attack removed a significant threat with apparently no loss of life (though the spread of the worm through networks resulted in monetary losses). Perhaps in this case it makes sense to think of the possibility that Stuxnet was actually consistent with cyber peace. (See also Brandon Valeriano and Benjamin Jensen's assessment of the potential de-escalatory function of cyber operations in Chapter 4 of this volume.)

But is it possible to thread that needle – to use low-intensity, carefully targeted cyber operations (limiting their harmful consequences) to avoid more hostile interventions – as a matter of strategy? And if so, do such actions promote cyber peace? The 2018 United States Department of Defense cyber strategy tries to do this with its "defend forward" approach to cyber security, and by "continuously engaging" adversaries (United States Cyber Command, 2018, pp. 4, 6). The implicit analogy to nuclear deterrence likely conditions decision makers' expectations, in my view. As Jason Healey explains, proponents of the strategy seek stability through aggressiveness. They assert that "over time adversaries will scale back the aggression and intensity of their operations in the face of US strength, robustly and persistently applied" (2019, p. 2). But Healey is cautious – noting the risk of negative outcomes – as persistent engagement could produce an escalatory cycle. In short, further characterizing the nature of cyber peace requires achieving greater clarity in differentiating between the kinds of cyber aggression that promote more peaceful outcomes rather than less.

The second point of distinction creates a boundary between problems that involve criminal violations versus those that rise to the level of aggressive breaches of cyber peace. Unlike cyber aggression, cybercrime, I suggest, is not the opposite of cyber peace. The scams, frauds, thefts, revenge porn postings, and pirated software that are everyday cybercrimes seem to me to be very bad sorts of things, but as policy problems they generally fall into the category of not lawful, rather than not peaceful. A society can be peaceful or cyber peaceful even in the presence of some crime; all societies have at least some crime. Countering cybercrime requires cyber law enforcement and international collaboration to deal with transnational crimes. Countering cyber aggression requires efforts toward (re)building cyber peace. These might include diplomacy, deterrence, or – the less peaceful alternative – aggression in return. Automatically folding cybercrime into the category of things that threaten cyber peace risks diluting the meaningfulness of cyber peace.

A caveat must be added, however. The boundary between cybercrime and cyber aggression is complicated by what Marietje Schaake describes as "the ease with which malign actors *with geopolitical or criminal goals* can take advantage of vulnerabilities across the digital world" (2020, emphasis added). The "or" should be understood as inclusive: "and/or." Cybercrimes can be used to attain geopolitical goals (acts of cyber aggression), criminal goals, or both. The 2017 "WannaCry" ransomware attack, attributed to North Korea, provides an example of both cyber

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

aggression and cybercrime. Initially, WannaCry was assumed to be the work of an ordinary criminal, but once North Korea's involvement became apparent, the evident geopolitical aim and the attack's aggressiveness became more important. We would sort WannaCry and similar aggressive actions in the category of “threats” to cyber peace rather than into the category of (only) “not lawful.”

Yet cybercrimes can, paradoxically, be tools *for* cyber peace too. Cybercrimes involving activities in support of human rights provide oppressed individuals and groups opportunities to fight back against their oppressors. Circumventing repressive surveillance technology might be an example of this. In that case, breaking the law could, arguably, be an example of cyber peace rather than a difference from it.

Third, the cyber peace pillar on multistakeholder collaboration assumes a distinction between cyber peace and non-cyber peace in terms of forms of governance. The definition of cyber peace includes a strong preference for developing “governance mechanisms by fostering multistakeholder collaboration” (Shackelford, 2016). Shackelford sees bottom-up multistakeholder governance as a form of polycentricity and as good in itself. But both polycentricity and multistakeholderism are problematic points of distinction for what is or is not cyber peaceful. Michael McGinnis and Elinor Ostrom (2012, p. 17), commenting on a classic article by Vincent Ostrom, Charles Tiebout, and Robert Warren (1961), call attention to how the authors:

> […] did not presume that all polycentric systems were automatically efficient or fair, and they never denied the fundamentally political nature of polycentric governance. The key point was that, within such a system, there would be many opportunities for citizens and officials to negotiate solutions suited to the distinct problems faced by each community.

A multistakeholder form of polycentric governance, however, involves not just citizens and officials negotiating solutions, but firms and other private actors as well, which potentially skews that political nature because the resources the different stakeholders have to draw upon in their negotiations can differ by orders of magnitude. As Michael McGinnis, Elizabeth Baldwin, and Andreas Thiel (2020) explain, polycentric governance can come to suffer from dysfunction because of structural forms that allow some groups to have outsized control over decision-making processes. And this is certainly true for a cyberspace governance organization like the Internet Corporation for Assigned Names and Numbers (ICANN), where the industry interests have significantly more say in outcomes than users. Furthermore, whereas polycentric governance evolves organically out of efforts to solve problems of different but related sorts, multistakeholderism is designed into the governance plan from its initiation, as was clearly the case with ICANN.

Moreover, as Kavi Joseph Abraham (2017) explains, stakeholderism is actually not about creating better forms of democratic governance. Rather, its origin story can be traced to “systems thinking” in engineering and related management

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

practices that emphasized the need for control of complexity. Complex systems, as engineers came to understand, involved multiple inputs, feedback loops, contingencies, outputs, etc. Controlling such systems required coordination of *all* those factors. That idea of coordinating all inputs into processes spilled over into the academic field of business management, where the firm came to be seen as a complex system. Control involved the coordination of material inputs plus the coordinated activity and decision-making of people – workers, managers, customers, shareholders, suppliers, communities affected by effluents from the firm's factory, etc. Groups that had a role to play were thus identified as "stakeholders," but unlike the assumed equality of citizens in a democracy, there was never any assumption that stakeholders should be equal or equivalent. Managing is about dealing with complexity, not about governing while protecting rights. We should not assume that multistakeholderism is uniquely suited to be the governance form for cyber peace.

8 METAPHORS

Finally, I raise the issue of metaphors and how they enable and limit thinking in some way (Cienki & Yanow, 2013; Lakoff & Johnson, 1980). First, is cyber peace the right metaphor that describes the sought-after goal? How would cyber peace be different from cyber order, cyber community, or cyber health? Given that much of the activity that goes on in cyberspace is commercial and given that commercial transactions are generally competitive rather than peaceful, does it make sense to talk about cyber *peace* when the goal is not friendly relations but, rather, a competitive market in which exchange can happen without the disruption of crime? How is cyber peace distinct from a well-functioning cyber market? Yet another alternative would be to rethink the marketization of cyberspace and to imagine instead a regulated utility and the provision of cyber services to the global public.

Moreover, by invoking peace in the context of what is often intended to be best practices of cyber security to maintain a stable Internet we fall prey to "inadvertent complicity" (Alker, 1999, p. 3), distracting attention from real violence. Overusing the peace metaphor flattens the differences between deeply consequential and ethically crucial peacemaking in the world, and getting people to use better passwords. We can see this flattening dynamic even when considering initiatives promising to save lives (anti-cyberbullying initiatives as a cyber peace practice, for example). I think cyberbullying is truly awful, and in the United States, it is a crime. It is often also a mental health challenge, both for the bully and bullied. It's a social pathology and a behavioral problem. It is also a cyber governance issue, as E. Nicole Thornton and I discussed in an article on the difficulties faced by owners of social media websites trying to prevent hijacking of their sites by bullies (Marlin-Bennett & Thornton, 2012). But is it useful to think of cyberbullying as a violation of cyber peace? (And doesn't doing so give the bully too much power?) Cyber peace becomes

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

hyperbole, notwithstanding the well-meaning campaigns such as that of the Cyber Peace Foundation (CyberPeace Corps, 2018). Peace is a strong word. By invoking peace (and war by implication), context and historicity can be washed away, obscuring the difference between cyberbullying and Russian cyber election disruptions that threaten to do grave harm to democracies.

9 A FINAL THOUGHT

In the oft-cited special issue of *International Organization* on international regimes, the final article was written by Susan Strange (1982). The title was "*Cave! hic dragones*: a critique of regime analysis." A note in smaller type at the bottom of the page reads "The title translates as 'Beware! here be dragons!' -an inscription often found on pre-Columbian maps of the world beyond Europe." The article, she explains in the first paragraph, does not ask "what makes regimes and how they affect behavior, it seeks to raise more fundamental questions about the questions." Her intent, instead, was to ask whether the regime concept is at all a useful advance for international political economy and world politics scholarship. She famously decided that the concept of the international regime was a bad idea for seven reasons (five main and two indirect). She was wrong. The concept of the international regime has endured and is widely accepted, and it has been useful. But I do not think that the concept of an international regime would have been nearly as well integrated into our scholarly lexicon now if it had not been for Strange's intervention. Over the subsequent years, proponents of the regime concept had to work to improve the concept to counter her claims, which were really quite fair, if expressed bluntly.

I do not have as negative an opinion of cyber peace as Strange did of international regimes, but her charge that the concept of international regimes was "imprecise and woolly" seems to fit the concept of cyber peace, as well. By analyzing the different meanings ascribed to cyber peace, I hope to do what Strange, intentionally or not, did for regimes theory: Make it better.

REFERENCES

Abbott, A. (1995). Things of Boundaries. *Social Research*, 62(4), 857–882.

Abraham, K. J. (2017). *Governing through Stakeholders: Systems Thinking and the Making of Participatory Global Governance* [Thesis, Johns Hopkins University].

Adler, E., & Pouliot, V. (2011). *International Practices*. Cambridge University Press.

Alker, H. R. (1999). *Ontological Reflections on Peace and War.* SFI Working Paper #99-02-011 (Unpublished Material). www.santafe.edu/research/results/working-papers/ontological-reflections-on-peace-and-war

Alker, H. R., & Greenberg, W. J. (1977). On Simulating Collective Security Regime Alternatives. In G. M. Bonham & M. J. Shapiro (Eds.), *Thought and Action in Foreign Policy* (pp. 263–305). Birkhäuser. https://doi.org/10.1007/978-3-0348-5872-4_9

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Alter, K. J., & Raustiala, K. (2018). The Rise of International Regime Complexity. *Annual Review of Law and Social Science*, *14*(1), 329–349. https://doi.org/10.1146/annurev-lawsocsci-101317-030830

Arias Sánchez, Ó. (2008). Foreword by the Patron of the Global Cybersecurity Agenda. In *Cybersecurity for ALL: ITU's Work for a Safer World* (pp. 4–5). International Telecommunications Union. www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-CYBER-2007-PDF-E.pdf

Balzacq, T. (2005). The Three Faces of Securitization: Political Agency, Audience and Context. *European Journal of International Relations*, *11*(2), 171–201. https://doi.org/10.1177/1354066105052960

Bigo, D. (2011). Pierre Bourdieu and International Relations: Power of Practices, Practices of Power. *International Political Sociology*, *5*(3), 225–258.

Buzan, B. (1993). From International System to International Society: Structural Realism and Regime Theory Meet the English School. *International Organization*, *47*(3), 327–352. https://doi.org/10.1017/S0020818300027983

Cienki, A., & Yanow, D. (2013). Why Metaphor and Other Tropes? Linguistic Approaches to Analysing Policies and the Political. *Journal of International Relations and Development*, *16*(2), 167–176. https://doi.org/10.1057/jird.2012.28

Cornut, J. (2015, December 1). *The Practice Turn in International Relations Theory*. Oxford Research Encyclopedia of International Studies. https://doi.org/10.1093/acrefore/9780190846626.013.113

CyberPeace Corps. (2018, May 23). Cyberbullying Is a Form of Bullying, and Adults Should Take the Same Approach to Address It: Support the Child Being Bullied, Address the Bullying Behavior of a Participant, and Show Children That #cyberbullying Is Taken Seriously. #CyberPeaceFoundation #CyberPeaceCorps https://t.co/bdbcE1Vd2l / Twitter [Social Media]. Twitter. https://twitter.com/cyberpeacecorps/status/999215740816429056?lang=ca

CyberPeace Institute – About Us. (2020). CyberPeace Institute. https://cyberpeaceinstitute.org/about-us Accessed June 14, 2021.

Diehl, P. F. (2016). Exploring Peace: Looking Beyond War and Negative Peace. *International Studies Quarterly*, *60*(1), 1–10. https://doi.org/10.1093/isq/sqw005

Diehl, P. F. (2019). Peace: A Conceptual Survey. In Oxford Research Encyclopedia of International Studies. Oxford University Press. https://doi.org/10.1093/acrefore/9780190846626.013.515

Dunn Cavelty, M. (2013). From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse. *International Studies Review*, *15*(1), 105–122. https://doi.org/10.1111/misr.12023

Foucault, M. (1986). Of Other Spaces. *Diacritics*, *16*(1), 22–27.

Gibson, W. (1994). *Neuromancer*. Ace Books.

Hansen, L. (2000). The Little Mermaid's Silent Security Dilemma and the Absence of Gender in the Copenhagen School. *Millennium*, *29*(2), 285–306. https://doi.org/10.1177/03058298000290020501

Healey, J. (2019). The Implications of Persistent (and Permanent) Engagement in Cyberspace. *Journal of Cybersecurity*, *5*(1), 1–15. https://doi.org/10.1093/cybsec/tyz008

Hobbes, T. (1651). *Leviathan*. Andrew Crooke, at the Green Dragon in St. Paul's Churchyard. www.gutenberg.org/files/3207/3207-h/3207-h.htm

Hofweber, T. (2005). A Puzzle About Ontology. *Noûs*, *39*(2), 256–283.

Krasner, S. D. (Ed.). (1982). International Regimes (special issue). *International Organization*, *36*(2).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Long, D., & Woolley, F. (2009). Global Public Goods: Critique of a UN Discourse. *Global Governance*, 15(1), 107–122.

Lyon, D. (2015). Beyond Cyberspace: Digital Dreams and Social Bodies. *Information Technology, Education, and Society*, 1(2), 5–21. https://doi.org/10.7459/ites/16.1.02

Marlin-Bennett, R., & Thornton, E. N. (2012). Governance within Social Media Websites: Ruling New Frontiers. *Telecommunications Policy*, 36(6), 493–501. https://doi.org/10.1016/j.telpol.2012.01.002

McGinnis, M. D., Baldwin, E. B., & Thiel, A. (2020). *When Is Polycentric Governance Sustainable? Using Institutional Theory to Identify Endogenous Drivers of Dysfunctional Dynamics*. https://ostromworkshop.indiana.edu/events/colloquium-series/index.html

McGinnis, M. D., & Ostrom, E. (2012). Reflections on Vincent Ostrom, Public Administration, and Polycentricity. *Public Administration Review*, 72(1), 15–25.

Ostrom, V., Tiebout, C. M., & Warren, R. (1961). The Organization of Government in Metropolitan Areas: A Theoretical Inquiry. *The American Political Science Review*, 55(4), 831–842.

Paris, R. (2001). Human Security: Paradigm Shift or Hot Air? *International Security*, 26(2), 87–102.

Parker, N., & Adler-Nissen, R. (2012). Picking and Choosing the 'Sovereign' Border: A Theory of Changing State Bordering Practices. *Geopolitics*, 17(4), 773–796. https://doi.org/10.1080/14650045.2012.660582

Permanent Monitoring Panel on Information Security of the World Federation of Scientists (WFS). (2009). *Erice Declaration on Principles for Cyber Stability and Cyber Peace*. World Federation of Scientists. www.aps.org/units/fip/newsletters/201109/barletta.cfm

Piñuelas, E. (2008). Cyber-Heterotopia: Figurations of Space and Subjectivity in the Virtual Domain. *Watermark*, 2, 152–169.

Pouliot, V., & Cornut, J. (2015). Practice Theory and the Study of Diplomacy: A Research Agenda. *Cooperation and Conflict*, 50(3), 297–315. https://doi.org/10.1177/0010836715574913

Reid, H. G., & Yanarella, E. J. (1976). Toward a Critical Theory of Peace Research in the United States: The Search for an "Intelligible Core." *Journal of Peace Research*, 13(4), 315–341. https://doi.org/10.1177/002234337601300404

Ruzicka, J. (2019). Failed Securitization: Why It Matters. *Polity*, 51(2), 365–377. https://doi.org/10.1086/702213

Schaake, M. (2020). The Lawless Realm. *Foreign Affiairs*, 99(6). www.foreignaffairs.com/articles/world/2020-10-13/lawless-realm

Searle, J. R. (2006). Social Ontology: Some Basic Principles. *Anthropological Theory*, 6(1), 12–29.

Shackelford, S. J. (2014). *Managing Cyber Attacks in International Law, Business, and Relations: In Search of Cyber Peace*. Cambridge University Press. https://doi.org/10.1017/CBO9781139021838

Shackelford, S. J. (2016). Business and Cyber Peace: We Need You! *Business Horizons*, 59(5), 539–548. https://doi.org/10.1016/j.bushor.2016.03.015

Shackelford, S. J., & Craig, A. N. (2014). Beyond the New Digital Divide: Analyzing the Evolving Role of National Governments in Internet Governance and Enhancing Cybersecurity Symposium. *Stanford Journal of International Law*, 50(1), 119–184.

Sjöstedt, R. (2017). *Securitization Theory and Foreign Policy Analysis*. Oxford Research Encyclopedia of Politics. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.479

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Strange, S. (1982). Cave! Hic Dragones: A Critique of Regime Analysis. *International Organization*, 36(2, International Regimes), 479–496.

Tewfik, N. (2010, November 26). Cyber Peace Initiative: Egypt's e-Safety Profile – "One Step Further Towards a Safer Online Environment". www.itu.int/dms_pub/itu-d/md/10/wtim8/c/D10-WTIM8-C-0036!!PDF-E.pdf

Touré, H. I. (2011). Cyberspace and the Threat of Cyberwar. In H. I.Touré & Permanent Monitoring Panel on Information Security of the World Federation of Scientists (WFS) (Eds.), *Cyberspace and the Threat of Cyberwar* (pp. 7–13). International Telecommunications Union. www.itu.int/pub/S-GEN-WFS.01-1-2011

Touré, H. I., & Permanent Monitoring Panel on Information Security of the World Federation of Scientists (WFS) (Eds.). (2011). *The Quest for Cyber Peace*. International Telecommunications Union. www.itu.int/pub/S-GEN-WFS.01-1-2011

United Nations Development Program. (1994). Human Development Report. http://hdr.undp.org/en/content/human-development-report-1994

United States Cyber Command. (2018). *Achieve and Maintain Cyberspace Superiority: Command Vision for US Cyber Command*. www.cybercom.mil/Portals/56/Documents/USCYBERCOM%20Vision%20April%202018.pdf?ver=2018-06-14-152556-010

Virtue Foundation Institute for Innovation and Philanthropy. (n.d.). Egypt's Cyber Peace Initiative. Virtue Foundation. https://virtuefoundation.org/project/cyber-peace-initiative/

Waever, O. (1996). European Security Identities. *Journal of Common Market Studies*, 34(1), 103.

Wegener, H. (2011). A Concept of Peace. In H. I.Touré & Permanent Monitoring Panel on Information Security of the World Federation of Scientists (WFS) (Eds.), *The Quest for Cyber Peace* (pp. 77–85). International Telecommunications Union. www.itu.int/pub/S-GEN-WFS.01-1-2011

Wiener, N. (1988). *The Human Use of Human Beings: Cybernetics and Society*. Da Capo Press.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

2

Domestic Digital Repression and Cyber Peace

Jessica Steinberg, Cyanne E. Loyle, and Federica Carugati

INTRODUCTION

The Chinese government has reportedly detained over a million Muslims in the northwestern region of Xinjiang (Maizland, 2019). The detainees, predominantly of the Uighur ethnic group, are being held in reeducation camps where they are forced to pledge loyalty to the Communist Party of China, renounce Islam, and learn Mandarin (Maizland, 2019). Officials in China purport that these camps are not only used for vocational training, but also cite the need to quell the influence of violent extremism in the Xinjiang population (Maizland, 2019). There are reports of prison-like conditions in these camps, including extensive surveillance, torture (Wen & Auyezov, 2018), and even forced sterilization (Associated Press, 2020). The Uighur population is also under extensive surveillance outside of these detention facilities. Alleged monitoring has included location surveillance through messaging apps such as WeChat, facial recognition technology used at police checkpoints, as well as biometric monitoring (Cockerell, 2019). These technologies are being used by the Communist Party as new digital tools for monitoring and controlling populations deemed threatening to the Chinese state.

Modern digital information and telecommunication technologies (ICTs) have changed the ways in which states and their citizens interact on a variety of fronts, including the provision of goods and services and the production of information and misinformation. As the case of the Uighur population in China suggests, ICTs have also changed the ways in which states address threats from their population. While these kinds of overt, blatant abuses carried out by authoritarian states against ethnic or religious minorities tend to capture much attention, the use of digital technologies for repression is by no means limited to authoritarian states (Dragu & Lupu, 2020). New technologies are shifting the ways in which all states, democratic as well as authoritarian, repress.

While improvements in technology have often been associated with liberation, digital technologies in the hands of governments willing to repress can be a major threat to respect for human rights and freedoms worldwide and, as such, they are

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

a danger to cyber peace. As defined in the Introduction to this volume, a *positive* cyber peace necessitates respect for human rights and freedoms and the spread of Internet access. These characteristics are threatened by domestic digital repression, which often includes intentionally limiting access to the Internet and cellular communications, and can both constitute and facilitate violations of human rights and freedoms. Our chapter focuses on the changing nature of repression through digital technologies as a risk to cyber peace. Differing from other contributions to this volume (see Chenou & Aranzales, Chapter 5), we explore the domestic side of the interaction between digital technologies and cyber peace. Digital technologies are transforming repression, but we still know very little about this transformation and its long-term impact on state behavior. We believe, however, that understanding the ways in which these technologies are reshaping state power and its relationship to its citizens is necessary to build a more peaceful and freer digital *and* analog world.

In this chapter, we provide a conceptual map of the ways in which ICTs impact state repression. This mapping exercise seeks to identify some initial sites of influence in order to further theorize and empirically evaluate the effects of ICTs on our current understandings of state repression. We begin by outlining a conceptual definition of digital repression informed by the extant literature on state repression. We then derive four constituent components of state repression and trace the impact of ICTs on each of our four components.[1] In conclusion, we discuss how our findings may inform or upend existing theories in the study of state repressive behavior.

1 REPRESSION AND DIGITAL REPRESSION

State repression refers to the actual or threatened use of physical violence against an individual or organization within a state for the purpose of imposing costs on the target and deterring specific activities believed or perceived to be threatening to the government (Goldstein, 1978, p. xxvii). Traditional modes of repression have been conceptualized based on their impact on the physical integrity of groups or individuals, or as restrictions on individual or group civil liberties. Physical integrity violations refer to violations of a person's physical being such as enforced disappearances, torture, or extrajudicial killings. Civil liberties violations include restrictions on press freedoms and information, and freedoms of association, movement, or religious practice.

All states repress, albeit in different ways and for different reasons (Davenport, 2007). Most scholars of state repression view the decision to repress as a rational calculus taken by political authorities when the costs of repression are weighed against its potential benefits (e.g., Dahl, 1966; Goldstein, 1978; Davenport, 2005). When

[1] We note here, but only in passing, that for both authoritarian and democratic governments, the relations with private ICT companies further complicate the strategic calculus. We address this issue below.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

the benefits of repression outweigh the costs, then states are likely to repress. The expected benefits of using repression are “the elimination of the threat confronted and the increased chance of political survival for leaders, policies, and existing political-economic relations” (Davenport, 2005, p. 122). In addition, repression may demonstrate strength and deter subsequent threats. Traditional costs of repression, on the other end of the equation, include logistical and monetary costs, as well as potential political costs. The literature on the dissent–repression nexus suggests that while repression may neutralize a threat in the short term, it has the potential to yield to more dissent in the longer term because of a backlash effect to state policies (Rasler, 1996; Koopmans, 1997; Moore, 1998; Carey, 2006). Democratic leaders who use particularly violent forms of repression may be penalized by voters (Davenport, 2005). Furthermore, leaders may suffer external political costs; for example, the international community may sanction leaders for excessive use of force against their civilian populations, or for behaviors that violate international human rights norms (Nielsen, 2013).

The advent of modern digital technologies has ushered in new forms of *digital* repression. Digital repression is the “coercive use of information and communication technologies by the state to exert control over potential and existing challenges and challengers” (Shackelford et al., in this volume, Introduction). Digital repression includes a range of tactics through which states use digital technologies to monitor and restrict the actions of their citizens. These tactics include, but are not limited to, digital surveillance, advanced biometric monitoring, misinformation campaigns, and state-based hacking (Feldstein, 2019). Modes of digital repression map onto the two modes of traditional repression mentioned above, physical integrity and civil liberties violations. Digital repression, while not directly a physical integrity violation, can facilitate or lead to such types of violations. For example, the data gathered by the Chinese state about the Uighur population has aided the government in locating and physically detaining large numbers of Uighurs. Digital repression can constitute both a civil liberties violation in and of itself, and facilitate the violation of civil liberties. For example, by limiting individual access to information and communication, the state violates the rights of citizens to access information. Alternatively, by closely monitoring the digital communications of social movements, states can deter or more easily break up political gatherings and protests. While states regularly gather and rely upon information about their citizens to conduct the work of governing, *digital repression* entails the use of that information for coercive control over individuals or groups that the state perceives as threatening.

As with traditional forms of repression, the use of digital repression can be seen in terms of a cost–benefit calculus on the part of the state. Yet, in the case of digital repression, this calculus is not well understood. Digital technologies impact the ways in which states identify and respond to threats, as well as the resources needed to do so. New technologies also impact the ways in which challengers, citizens, and the international community will experience and respond to the state’s behavior,

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

in turn affecting the costs and benefits of using digital repressive strategies. For example, the costs of digitally monitoring social movement participation through social media may have large upstart costs in terms of infrastructure and expertise. Yet, those initial costs may be offset over future threats. In certain circumstances, digital repression may reduce audience costs associated with traditional forms of repression[2] as these newer forms of repressive behavior may be easier to disguise. Alternatively, if digital repression is hidden from the public, it may be less likely to deter future threats, as challengers may not fully understand the levels of risk involved in challenging the state. In sum, it is likely that digital repression is shifting the cost–benefit analysis of state repression. However, we have yet to adequately theorize how this analysis might differ from, and relate to, a cost–benefit analysis of the use of traditional state repression.

Before we map how ICTs are reshaping state repression, we first place some scope conditions on the set of technologies that are relevant for our inquiry. Within the last decade, scholars have begun to develop frameworks and explore empirical patterns related to digital repression in a still nascent literature.[3] This work has examined a wide range of technologies, strategies, and platforms, including Internet outages (Howard et al., 2011), social media use (Gohdes, 2015), and surveillance technologies (Qiang, 2019). Building on this work, we focus on the technological developments that facilitate two kinds of activities: (1) access to new and potentially diverse sources of information and (2) near instantaneous communication among individual users. While neither of these activities is a fundamentally new use of technology, the volume of information available, the number of individuals that can access and communicate information, and the speed at which exchanges can occur are new. Therefore, we are interested in ICTs that combine cellular technology, the Internet and its infrastructure, the software and algorithms that allow for large-scale data processing, and the devices that facilitate access to the Internet (e.g., computers and smart phones).[4]

As Shackelford and Kastelic (2015) detail, as states have sought to protect critical national infrastructure from cyber threats, they have pursued more comprehensive state-centric strategies for governing the Internet. This has led to the creation of national agencies and organizations whose purpose is to monitor communication and gather data and information about foreign as well as domestic ICT users. This is true for both democracies and autocracies. But whereas we tend to associate democracies with robust legal protections and strong oversight institutions (especially with

[2] Traditional forms of repression impose political costs on the leader when a variety of groups (both domestic and international) observe these forms of repression and respond in ways that penalize the leader.

[3] For example, Deibert et al. (2008); Howard et al. (2011); Dainotti et al. (2011); Gohdes (2015); Rydzak (2015); Hellmeier (2016); Wagner (2018); Deibert (2019); Qiang (2019); and Diamond (2019).

[4] While cellular technology is certainly not new, the widespread use of smartphones allows citizens to make use of cellular technology to access the Internet.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

respect to the private sector), we do not need to travel far to find cases of democracies with timid approaches to oversight and protection of individual rights – the obvious example is the US government's reluctance to reign in technology giants such as Apple or Facebook. Once governments gather information about users, they can engage in two kinds of activities that may lead to violations of citizen's rights through either physical integrity or civil liberties violations. First, states can monitor and surveil perceived existing or potential threats. Second, states can limit access to ICTs or specific ICT content for individuals or groups perceived to present a threat to the state. The monitoring of threats and restrictions on threatening behavior are not new behaviors for states. However, digital technologies provide new opportunities for states to exercise control.

2 THE IMPACT OF ICTs ON STATE REPRESSION

We argue that state repression requires four specific components to function effectively. First, a state must have the ability to *identify a threat*. Second, the state must have the *tactical expertise* to address the threat. Third, a state must be able to compel *responsive repressive agents* to address a threat in a specified way. And fourth, the state must have a physical *infrastructure of repression* that facilitates addressing the threat, or at least does not make addressing the threat prohibitively costly. Below we discuss these components of repression and conceptualize the potential impacts of ICTs on each.

2.1 *Threat Identification*

Governments engage in repression in order to prevent or respond to existing or potential threats. The first component of repression, therefore, requires the state to be able to effectively identify and monitor these threats. Identifying and monitoring threats is costly. These costs are largely associated with gathering information which, depending on the nature of the threat, are likely to vary. Costs vary depending on whether the government is responding to an existing and observable threat (such as a protest or riot, or formal political opposition) or whether the government is attempting to detect a potential threat, which could be more difficult to identify.

Threat identification requires that governments have cultural, linguistic, and geographic knowledge (Lyall, 2010). The costs associated with gathering this kind of knowledge vary depending on context (Sullivan, 2012). For example, there are urban/rural dynamics when it comes to threat identification. In some circumstances, it is easier for the state to monitor threats in an urban center, which may be close to the political capital, rather than in the hinterland, where geographic barriers could hinder information collection (Herbst, 2000). Conversely, in other contexts, urban concentrations may make it more costly to identify and isolate a particular threat. The size of a potential threat also impacts the costs of threat identification.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Mass surveillance of the Uighur population, for example, requires the identification and monitoring of approximately twelve million people.

In both traditional forms of state repression and digital repression, information is central to identify existing and potential challenges to the state. Digital technologies offer the possibility of significantly lowering the costs of information collection for the state. The speed and volume with which information can be collected and processed is far greater than with any monitoring or surveillance techniques of the past. Moreover, as Deibert and Rohozinski (2010, p. 44) write, "Digital information can be easily tracked and traced, and then tied to specific individuals who themselves can be mapped in space and time with a degree of sophistication that would make the greatest tyrants of days past envious." Individuals leave digital footprints, online or through cellular communication, with information that ties them to specific beliefs, behaviors, and locations. States can also track a much broader section of the population than was ever previously possible. For example, states threatened by mass mobilization can now closely monitor, in real time, crowd formations with the potential to become mass rallies, allowing police to be put on standby to immediately break up a protest before it grows (Feldstein, 2019, p. 43).

The availability of less overt forms of threat detection may open up new strategic possibilities for governments, shaping their choice among forms of digital repression as well as between digital and traditional repression. For example, it is possible that a state would refrain from using certain monitoring tactics that are visible and attributable to the state, not because they would be useless in identifying a particular threat, but because the government does not want to tip its hand about its repressive capacity. In this circumstance, a government might choose to monitor a population, for example, rather than engage in mass incarceration. Still, digital technologies for threat identification also carry costs. The Xinjian authorities, for example, reportedly budgeted more than $1 billion in the first quarter of 2017 for the monitoring and detention of the Uighur population there (Chin & Bürge, 2017). However, this figure is likely lower than the amount the Chinese state would have spent to construct a comparable system without using digital technologies (Feldstein, 2019, pp. 45–46). Furthermore, once those investments have been made, a form of path dependence is likely to ensure that the new expertise will continue to lead to particular forms or repression (as we discuss in the next section on tactical expertise).

The ability to access more, indeed enormous, amounts of information has the potential to increase the cost of threat identification. In fact, such volume of searchable data raises the challenge of identifying a threatening signal in an ever growing pile of digital noise. The problem, then, is not simply finding a signal, but the possibility that more digital noise could result in biased or wrong signals. Digital surveillance is often a blunter monitoring tool than individual surveillance techniques of the past, given the quantity of digital information which is now available. However, the development of algorithms and reliance on artificial intelligence for sifting through large amounts of information can significantly lower threat identification

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

costs for states. But such tools, in turn, require new forms of tactical expertise. We discuss this issue in the next section.

2.2 *Tactical Expertise*

Once a threat has been identified, in order to repress effectively the state must have the ability to address the threat. Tactical expertise refers to the actual know-how of repression – that is, the skillsets developed by the state to exert control, ranging from surveillance techniques to torture tactics. A number of studies demonstrate that repressive tactics are both taught and learned (see, e.g., Rejali, 2007). Understanding the tactical expertise of a state when it comes to repression can tell us not only about the ability of the state to repress in the first place, but also about the type of repression the state is most likely to engage in when faced with a particular threat. Each state will have a specific skillset that enables it to repress in certain ways, but not in others. Certain techniques of repression may be unavailable to a state, or they would require the costs of appropriating a new skill. States may or may not be able to incur those costs. For example, states may invest in becoming experts at torture or, instead, they might invest in tools of riot policing. The "coercive habituation" of a state suggests that, if the state has engaged in repression or a type of repression in the past, this lowers the costs of applying the same form of repression in the future (e.g., Hibbs, 1973; Poe & Tate, 1994; Davenport, 1995, 2005). Therefore, the likelihood of becoming proficient at a particular repressive tactic is (at least in part) a function of the state's history of threats and threat perception, as well as the history of the state's response to these threats. We expect states to have varying levels of expertise across a variety of coercive tactics. These levels of expertise are reflected in the training centers, organizational infrastructure, and command structure of particular governmental actors charged with implementing repressive tactics.

Digital repression requires *technological* knowhow or expertise. This might be reflected in the availability of experts trained in information technology, fixed network and mobile technologies, or critical systems infrastructure. Technological expertise ultimately corresponds to the country's reservoir of expert knowledge in the use, maintenance, and control of ICT systems. Given the resource requirements of acquiring this form of expertise in order to implement certain forms of digital repression, some governments may be unable or unwilling to incur the costs of developing the relevant skillset.

The relevant type and level of technological know-how required for digital repression varies based on who a state targets with digital repression, as well as what (if any) content is being restricted. Targets of digital repression can range from individual users to specific groups across specific geographic regions, or the whole country. States can also restrict access to, or publication of, information ranging from single websites to entire platforms or applications. In some cases, states can engage in a wholescale Internet or cellular communications blackout. Targeting individual users, as opposed to large swathes of the population, may be

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

more costly as it requires a higher level of threat identification, and potentially greater levels of technical and algorithmic expertise. Similarly, targeting a specific website or single platform is often more costly and requires greater technical capacity than enforcing a wholescale Internet blackout.[5] The presence of a "kill switch" in some countries means that the state can easily disrupt telecommunications by creating a network blackout, a crude though often effective form of digital repression. It is more technically difficult, for example, to restrict access to a specific platform such as WhatsApp or Twitter, or to block access for a specific individual or group, especially if targeted individuals have their own technological expertise to develop effective workarounds (i.e., virtual private networks, for example). The target and form of digital repression is therefore influenced by the state's availability and nature of tactical expertise.

2.3 *Responsive Repressive Agents*

Once a threat has been identified and a repressive strategy has been chosen, governments rely on repressive agents to implement that strategy. In general, the leader himself/herself is decidedly *not* the agent of repression. Instead, the state relies on a repressive apparatus. Unpacking the state into a principal (leader) and an agent (the security apparatus), as much of the literature on repression does, is helpful for demonstrating that organizational capacity and power are necessary dimensions of the state's ability to repress. This ability corresponds to the level of centralization, the degree of organization, and the level of loyalty of the repressive agents. The agents of repression are often confined within a set group of organizations that vary by regime and regime type, such as the police, military, presidential security, etc. On rare occasions, state repression can be outsourced to agents not directly under the command of the state, for example, pro-government militias, vigilante groups, or private military contractors. The outsourcing of repression further complicates the issue of ensuring compliance from repressive agents. The state must have the ability to develop these organizations as loyal, responsive agents endowed with the expertise to implement the relevant repressive tactic.

Some forms of digital repression may require fewer repressive agents, simplifying principal-agent issues for repressive states. For example, digital repression might be carried out by a few technical experts within a government agency, or by an automated algorithm. One intuitive possibility is that requiring fewer agents to carry out a repressive action is less costly because of lower coordination costs and gains in efficiency. In the past, mass surveillance required an extensive network of informers. For example, in Poland in 1981, at the height of the Sluzba Bezpieczenstwa's (Security Service) work to undermine the Solidarity

[5] These costs are also likely to vary depending on the website or platform, since many larger companies (Google, for example) have begun to develop their own Internet infrastructure.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

movement, there were an estimated 84,000 informers (Day, 2011). New technologies produce the same level of surveillance (or greater) from the work of far fewer people. However, while fewer agents may be easier to coordinate, failure or defection by one among only a few repressive agents may be more costly in comparison to failure by one among thousands.

Online communication and access to digital space further requires a telecommunication company or Internet provider which may be outside of the state's direct control. Though governments often have ownership stakes in these companies, which are seen as a public utility, the companies themselves remain independent actors. The level or ease of control that the state exhibits over the ICT sector varies, thereby shaping how easily the state can compel the sector to engage in repressive behaviors, such as monitoring usage or controlling access. For certain forms of digital repression, governments require greater capacity to compel specific actions on the part of these actors (such as shutting down the Internet, limiting access to specific platforms, limiting broadband access, etc.). The power to compel these actions is determined by government involvement in the sector and by market characteristics (industry structure and the number of actors), as well as existing legal protections – for example, regulations determining whether or when Internet service providers are required to turn over data to the state. In Europe, the General Data Protection Regulation, though aimed primarily at private actors, gives greater control to users over their individual data, and therefore makes it more difficult for governments to obtain access to personal, individual user data. If firms cannot collect it, they cannot be compelled to provide it to governments. In these ways, ICTs have the potential to simultaneously simplify and complicate the state's relationship to its repressive agents, making it difficult to anticipate how ICTs will change the costs of repression in this regard.

2.4 *Infrastructure of Repression*

The capacity to apply repressive pressure in response to an identified threat requires what might be called an *infrastructure of repression*. This infrastructure should be thought of as the physical, geographic, or network characteristics that make it more or less costly (in terms of effort and resources) to engage in repression. At its most basic, repression infrastructure refers to sites of repression, such as prisons and detention facilities. It also refers to the physical environment in which repressive tactics are executed, which include the man-made and natural terrain that shapes the costs of repression (Ortiz, 2007). In civil war literature, many have argued that the existence of a paved road network reduces the government costs of repressing a threat because government vehicles and soldiers can more easily access their targets (Buhaug & Rød, 2006). This result echoes James Scott's discussion of the rebuilding of Paris by Hausmann, which had the explicit goal of constructing a gridded road that government troops could use to more easily reach any part of the city to prevent or put down riots or protests in the aftermath of the French Revolution (Scott, 1998).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

The concept of a repressive infrastructure has an intuitive analogue in the digital repressive space due to the physicality of telecommunications. The technologies that facilitate communication and the diffusion and exchange of information require physical infrastructures – the cellular towers, the fiber-optic cables, the data centers, and interconnection exchanges[6] – that are the building blocks of the networks of digital and cellular communication.

Scholars have begun to use the characteristics of a country's Internet technology network of autonomous systems (ASs) and the number of "points of control" to rank and characterize digital infrastructure in terms of the level of control governments can exert over citizen access to telecommunications networks and the data flowing across them (Douzet et al., 2020). Autonomous systems route traffic to and from individual devices to the broader Internet, which in turn is a collection of other ASs. The AS is, therefore, the primary target of regulation, monitoring, and interference by the state. Because most ASs are part of a larger network of systems, the vast majority of Internet traffic flows through a relatively small number of ASs within a country (often between three and thirty).[7] The minimum set of ASs required to connect 90 percent of the IP addresses in a country are called "points of control." The more points of control there are in a country, the more costly it is to regulate or restrict digital communication (both in terms of skills and equipment).

Roberts et al. (2011) have mapped two characteristics of in-country networks: the number of IP addresses (a proxy for individual users) per point of control and the level of complexity of the network within a country (the average number of ASs a user has to go through to connect to the Internet). Countries with more centralized systems and fewer points of control are places in which governments can much more easily exert control over access to the Internet for large portions of the population, and over the data that travel across the network. For example, as of 2011, the Islamic Republic of Iran had only one single point of control, with over four million IP addresses and a low network complexity score, which ensured that the state could easily control the entire Internet. According to Roberts, "in Iran, shutting down each network takes only a handful of phone calls" (Roberts et al., 2011, p. 11). As a result, such systems may require less expertise, less time, and less equipment to obtain and collect data, monitor users, or limit access. The greater the level of control over the infrastructure a state commands, the lower the costs to engage in digital repression.[8]

[6] Also called Internet exchange points (IXP).

[7] There are an estimated four billion Internet users globally, each of whom must connect to the Internet through an AS (of which there are an estimated 60,000 in total).

[8] The infrastructure of digital repression is not, however, an entirely exogenous component of the state's decision to engage in a particular kind of repression. The nature of these network characteristics is not accidental, but often designed to facilitate state control. Referring back to the invention of the Internet, Roberts et al. (2011) note that "the birth of the Internet as the split of ARPANET into two politically distinct networks was an explicitly political decision – intended to allow distinct modes of political control over the distinct networks" (p. 3).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

In addition to the network characteristics of the Internet within a country, the infrastructure for digital repression is also characterized by how the majority of individuals communicate and access the Internet. Smart cellular phones are by far the most common devices used to access the Internet, in addition to facilitating voice and text communication. They provide an additional point in the digital infrastructure where states can exert control. For example, states may impose regulations requiring proof of identification in order to obtain a cell phone and sim card. By doing so, they are able to collect large amounts of data about who owns which devices, and thereby monitor individual communications and data (including locational data).

3 NEW THOUGHTS ON DIGITAL STATE REPRESSION AND CYBER PEACE

States repress when the benefits of repression outweigh its costs. But when states repress using digital tools, how does this calculus change? How do digital forms of repression coexist with, or substitute for traditional forms of repression? And how does the combination of traditional and digital forms of repression affect the goal of cyber peace?

These are some of the questions we need to address in order to tackle the complex interactions among domestic state repression, digital technologies, and cyber peace. This chapter does not provide comprehensive answers, but it begins to unpack these interactions. In particular, our contribution is twofold. First, we break down repression into four constituent components, facilitating a conceptualization of repressive actions that cuts across the traditional/digital divide. This framework provides a useful workhorse for advancing research on the empirical patterns of repression. Second, we use this mapping to begin to explore the complex ways in which digital repression can impact each of the four components.

The value of our mapping exercise for scholars and practitioners in the field of cyber peace and cybersecurity emerges most poignantly in the reflections we offer about the tradeoffs between domestic and international security. For example, an Internet architecture that has a single point of control allows for governments to easily control access to the Internet and monitor data traveling over the network, but it also presents a vulnerability to foreign actors who only need to obtain control of, or infiltrate that point of control in order to gain access to domestic networks. This was in fact the case with Iran which, as noted earlier, had a single point of control until 2011. However, in recognition of the potential vulnerabilities to foreign intrusions that this created, Iran has since sought to add complexity to its digital infrastructure (Salamatian et al., 2019). But, as a result, it also had to acquire greater expertise to manage this complexity, developing a broader range of tools to monitor users and control access (Kottasová and Mazloumsaki, 2019). Another issue concerns strategic interdependencies: States may need to rely on international collaborations to carry out repression within their own borders. This problem is

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

particularly acute given the fact that servers are often housed in data centers outside the country in which most of their users reside.

Our contribution also suggests that some of the core insights in the literature on repression need reconsideration. For example, the literature on repression suggests that while all regime types repress, democracies repress less than autocracies (Davenport, 2007). However, this may not be true in the case of digital repression. Given the importance of the audience costs that we tend to associate with democratic regimes, we might expect democracies to invest and engage more in forms of repression that are more difficult to detect and observe. Perhaps, more interestingly, the existing literature suggests that democracies and autocracies differ with respect to the way they use information, and such differences expose them to different threats (Farrell & Schneier, 2018). This difference may shape the cost–benefit analysis of engaging in certain forms of digital or traditional repression in distinct, regime-specific ways.

Moreover, the repression literature further suggests that under certain conditions state repression may increase rather than eliminate dissent. The dissent–repression nexus may require reexamination in light of how ICTs are reshaping repression. The addition of a new menu of repressive tactics that can be used in conjunction with, or in place of, traditional forms of repression may lead the state to more effectively mitigate or eliminate threats in ways that make them less likely to resurface or produce backlash. This is in part because of the addition of more covert forms of repression that might be less observable and generate fewer grievances down the line.

Finally, the literature suggests that repression requires high levels of state capacity. However, when states repress through computers, and not police and tanks, repression may rely on sectors and skills that we do not currently measure or think of as relevant dimensions of state capacity. In particular, taking a more granular, multidimensional approach to state capacity, with particular attention to the specific capacity to repress, may shed new light on the relationship between generalized state capacity for repression and state capacity for digital repression.

These observations also yield a distinct, methodological question: If digital repression makes preemptive repression more effective, how can we continue to effectively measure repression since we will have many more unobservable cases in which repression preempted the emergence of an observable threat? Although we do not venture to answer this question, we hope that our chapter offers a starting point for a comprehensive analysis of repression in its traditional as well as digital forms.

The ability of states to violate the political and civil liberties of their populations through digital technologies is a direct threat to cyber peace. While often overlooked in our more internationalized discussion of cyber warfare, how states use and misuse digital technologies to monitor and control their populations is a subject that requires much more attention both because it can shape and be shaped by internationalized cyber warfare, and also because it is an important empirical and normative concern in and of itself.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

REFERENCES

Associated Press (2020, June 29). China cuts Uighur births with IUDs, abortion, sterilization. The Associated Press. https://apnews.com/article/ap-top-news-international-news-weekend-reads-china-health-269b3de1af34e17c1941a514f78d764c

Buhaug, H., & Rød, J. K. (2006). Local determinants of African civil wars, 1970–2001. *Political Geography*, 25(3), 315–335.

Carey, S. C. (2006). The dynamic relationship between protest and repression. *Political Research Quarterly*, 59(1), 1–11.

Chin, J., & Bürge, C. (2017, December 17). Twelve days in Xinjiang: How China's surveillance state overwhelms daily life. *The Wall Street Journal*. www.wsj.com/articles/twelve-days-in-xinjiang-how-chinas-surveillance-state-overwhelms-daily-life-1513700355

Cockerell, I. (2019, May 9). Inside China's massive surveillance operation. *Wired*. www.wired.com/story/inside-chinas-massive-surveillance-operation/

Dahl, R. A., Ed. (1966). *Political oppositions in western democracies.* Yale University Press.

Dainotti, A., Squarcella, C., Aben, E., Claffy, K., Chiesa, M., Russo, M., & Pescapè, A. (2011, November). Analysis of country-wide internet outages caused by censorship [Manuscript]. Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. Berlin, Germany. https://doi.org/10.1145/2068816.2068818

Davenport, C. (1995). Multi-dimensional threat perception and state repression: An inquiry into why states apply negative sanctions. *American Journal of Political Science* 39(3), 683–713.

Davenport, C. (2005). Understanding covert repressive action: The case of the US Government against the Republic of New Africa. *Journal of Conflict Resolution* 49(1), 120–140.

Davenport, C. (2007). *State repression and the domestic democratic peace.* Cambridge University Press.

Day, M. (2011, October 18). Polish secret police: How and why the Poles spied on their own people. The Telegraph. www.telegraph.co.uk/news/worldnews/europe/poland/8831691/Polish-secret-police-how-and-why-the-Poles-spied-on-their-own-people.html

Deibert, R. (2019). The road to digital unfreedom: Three painful truths about social media. *Journal of Democracy*, 30(1), 25–39.

Deibert, R., Palfrey, J., Rohozinski, R., Zittrain, J., & Gross Stein, J. (2008). *Measuring global internet filtering*. MIT Press.

Deibert, R., & Rohozinski, R. (2010). Liberation vs. control: The future of cyberspace. *Journal of Democracy*, 21(4), 43–57.

Diamond, L. (2019). The road to digital unfreedom: The threat of postmodern totalitarianism. *Journal of Democracy*, 30(1), 20–24.

Douzet, F., Pétiniaud, L., Salamatian, L., Limonier, K., Salamatian, K., & Alchus, T. (2020, May 26–28). Measuring the fragmentation of the Internet: The case of the Border Gateway Protocol (BGP) during the Ukrainian crisis [Manuscript]. 12th International Conference on Cyber Conflict (CyCon). Tallinn, Estonia.

Dragu, T., & Lupu, Y. (2020). Digital authoritarianism and the future of human rights. *International Organization*. http://yonatanlupu.com/Dragu%20Lupu%20IO.pdf

Farrell, H. J., & Schneier, B. (2018, October). Common-knowledge attacks on democracy. Berkman Klein Center Research Publication. https://cyber.harvard.edu/story/2018-10/common-knowledge-attacks-democracy.

Feldstein, S. (2019). The road to digital unfreedom: How artificial intelligence is reshaping repression. *Journal of Democracy*, 30(1), 40–52.

Gohdes, A. (2015). Pulling the plug: Network disruptions and violence in civil conflict. *Journal of Peace Research*, 52(3), 352–367.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Gohdes, A. (2020). Repression technology: Internet accessibility and state violence. *American Journal of Political Science*, 64(3), 488–503.

Goldstein, R. J. (1978). *Political repression in modern America from 1870 to the present*. GK Hall & Company.

Hellmeier, S. (2016). The dictator's digital toolkit: Explaining variation in Internet filtering in authoritarian regimes. *Politics & Policy*, 44(6), 1158–1191.

Herbst, J. (2000). *States and power in Africa*. Princeton University Press.

Hibbs, D. A. (1973). *Mass political violence: A cross-national causal analysis*. Wiley.

Howard, P. N., Agarwal, S. D., & Hussain, M. M. (2011). When do states disconnect their digital networks? Regime responses to the political uses of social media. *The Communication Review*, 14(3), 216–232.

Koopmans, R. (1997). Dynamics of repression and mobilization: The German extreme right in the 1990s. *Mobilization: An International Quarterly*, 2(2), 149–164.

Kottasová, I., & Mazloumsaki, S. (2019, November 19). The "internet as we know it" is Off in Iran. Here's why this shutdown is different. *WRAL*. www.wral.com/the-internet-as-we-know-it-is-off-in-iran-heres-why-this-shutdown-is-different/18778492/?version=amp [Accessed: April 20, 2021].

Lyall, J. (2010). Are coethnics more effective counterinsurgents? Evidence from the second Chechen War. *American Political Science Review*, 104(01), 1–20.

Maizland, L. (2019, November 25). China's repression of Uighurs in Xinjiang. Council on Foreign Relations. www.cfr.org/backgrounder/chinas-repression-uighurs-xinjiang

Moore, W. H. (1998). Repression and dissent: Substitution, context, and timing. *American Journal of Political Science*, 42(3), 851–873.

Nielsen, R. A. (2013). Rewarding human rights? Selective aid sanctions against repressive states. *International Studies Quarterly*, 57(4), 791–803.

Ortiz, D. (2007). Confronting oppression with violence: Inequality, military infrastructure and dissident repression. *Mobilization*, 12(3), 219–238.

Poe, S., & Tate, C. N. (1994). Repression of personal integrity rights in the 1980s: A global analysis. *American Political Science Review* 88, 853–872.

Qiang, X. (2019). The road to digital unfreedom: President Xi's surveillance state. *Journal of Democracy*, 30(1), 53–67.

Rasler, K. (1996). Concessions, repression, and political protest in the Iranian revolution. *American Sociological Review*, 61(1), 132–152.

Rejali, D. (2007). *Torture and democracy*. Princeton University Press.

Roberts, H., Larochelle, D., Faris, R., & Palfrey, J. (2011). Mapping local internet control. In *Computer communications workshop (Hyannis, CA, 2011), IEEE*.

Rydzak, J. (2015). The digital dilemma in war and peace: The determinants of digital network shutdown in non-democracies [Manuscript]. International Studies Association 57th Annual Convention. Atlanta, GA, United States.

Salamatian, L., Douzet, F., Limonier, K., & Salamatian, K. (2019). The geopolitics behind the routes data travels: A case study of Iran. arXiv. https://arxiv.org/ftp/arxiv/papers/1911/1911.07723.pdf

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

Shackelford, S. J., & Kastelic, A. (2015). Toward a State-centric cyber peace: Analyzing the role of national cybersecurity strategies in enhancing global cybersecurity. *NYUJ Legis. & Pub. Pol'y* 18, 895.

Sullivan, C. M. (2012). Blood in the village: A local-level investigation of State Massacres. *Conflict Management and Peace Science*, 29(4), 373–396.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Wagner, B. (2018). Understanding Internet shutdowns: A case study from Pakistan. *International Journal of Communication*, 12(1), 3917–3938.

Wen, P., & Auyezov, O. (2018, November 29). Turning the Desert into Detention Camps. *Reuters*. www.reuters.com/investigates/special-report/muslims-camps-china/

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

PART II

Modalities: How Might Cyber Peace Be Achieved? What Practices and Processes Might Need to Be Followed in Order to Make It a Reality?

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

3

Information Sharing as a Critical Best Practice for the Sustainability of Cyber Peace

Deborah Housen-Couriel

1 INTRODUCTION: FRAMING THE RELATIONSHIP BETWEEN INFORMATION SHARING AND CYBER PEACE

The concept of cyber peace brings a much-needed, innovative perspective to discussions of the governance of cyberspace. The ambiguity, conflicting terminology, and lack of transparency with respect to activities by state and nonstate actors have characterized efforts to conceptualize and analyze this new area of human endeavor at least since John Perry Barlow's 1996 Declaration of the Independence of Cyberspace. Barlow's (1996) proclamation that claimed cyberspace as a home for the "civilization of the Mind" and a "global social space" that must be kept free of governments, state sovereignty, and legal constructs – in effect, exempt from any type of governance – marked early on in the life of online activities the challenges and tensions that remain today for the global collective action problem of cyberspace governance. Thus, the distinctive perspective of cyber peace has the potential to set our analytical sights anew and to provide a framework for moving ahead with the normative projects connected to the aspects of cyberspace governance, including the ongoing elucidation of binding rules of international and domestic law that are applicable to cyberspace activities of state and nonstate actors.

Building on previous chapters that treat the concept of cyber peace in depth, the following definition focuses on four specific elements:

> Cyber peace is [...] not [...] the absence of conflict [...]. Rather it is the construction of a network of multilevel regimes that promote global, just and sustainable cybersecurity by clarifying the rules of the road for companies and countries alike to help reduce the threats of cyber conflict, crime and espionage to levels comparable to other business and national security risks. To achieve this goal, a new approach to cybersecurity is needed that seeks out best practices from the public and private sectors to build robust, secure systems and couches cybersecurity within the larger debate on internet governance (Shackelford, 2014, pp. xxv–xxvi).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

The four elements emphasized in the above definition describe the fundamental connection between the goals of cyber peace and information sharing (IS), the subject of this chapter (Johnson et al., 2016, p. iii).[1] Clarification of "rules of the road," whether these are binding or voluntary; threat reduction, risk assessment, and best practices for carrying out these three tasks are precisely the substantive contribution that IS makes to the cybersecurity postures and strategies of stakeholders participating in any given IS platform. As detailed herein, such a platform optimally defines threshold norms of permissible and nonpermissible online behavior on the part of all actors, establishing the criteria for determining whether an individual, private organization, country, group of hackers, or even another autonomously acting computer has violated a rule (Deljoo et al., 2018, p. 1508). It also reduces vulnerability to cyber threats by lessening the informational asymmetries that characterize hostile cyber activities to the advantage of the attacker, and contributes to organizational risk assessment by integrating the information shared by other participants in the IS community into heightened "cyber situational awareness" for all sharers. Fourth, IS is readily framed and understood by a multiplicity of actors at the domestic level – private, governmental, and individual – as a *best practice* and, at the international level, as a *confidence-building measure* (CBM) for building trust among state and nonstate actors.[2] These two characterizations of IS in the domestic and international jurisdictional arenas, respectively, are evidenced by the inclusion of IS modalities in many instances of national law and policy, as well as tens of multilateral and bilateral instruments for governing cyberspace at the international level (Housen-Couriel, 2017, pp. 46–84). Five examples of the latter are the 2015 Shanghai Cooperation Organization's International Code of Conduct for Information Security, the UN GGE Report of July 2015, the OSCE's Confidence-Building Measures for Cyberspace of 2016, the EU's Network and Information Security Directive that entered into force in August 2016; and the 2018 Paris Call for Trust and Security in Cyberspace.

When IS implemented as a voluntary or recommended best practice or CBM in the context of these regulatory arrangements – rather than as a mandated regulatory requirement – it has the advantage of bypassing the legal challenges of achieving formal and substantive multistakeholder agreement on cyber norms. The difficulties

[1] The 2016 NIST Guide to Cyber Threat Information Sharing has noted the advantages of IS measures as a means of leveraging the collective knowledge, experience, and capabilities of both state and nonstate actors within the sharing community in order to enhance the capability of each to make informed decisions regarding development of policies, defensive capabilities, threat detection techniques, and mitigation strategies.

[2] On information sharing as an enabler of trust building to resolve collective action problems see, for example, Ostrom et al. (1990) ("By voluntarily sharing the costs of providing information – a public good – participants learned that it was possible to accomplish some joint objectives by voluntary, cooperative action."); and Ostrom et al. (2012), pp. 23, 79, 81–82, 88, and 93 (where IS constitutes an element of the Socio-Ecological System, or SES concept used by Elinor Ostrom to analyze ecosystems addressing a collective action problem).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

of such normative barriers are often observed as characteristic of the contemporary cyber lay of the land. Either as a best practice (at the domestic level) or a CBM at the international level, IS has the advantage of bypassing the present challenges of achieving formal and substantive multistakeholder agreement on cyber norms that are inherent elements of national and multilateral legal regimes for the governance of cyberspace (Macak, 2016; Ruhl et al., 2020).

We propose in this chapter that, as IS platforms provide increasingly relevant, timely, and actionable data on vulnerabilities, including zero-day vulnerabilities (Ablon & Bogart, 2017); adversaries' tactics, techniques, and procedures; malware tool configurations; and other tactical and strategic threat indicators, stakeholders will become more incentivized to increasingly trust IS platforms and to utilize them for both real-time response to hostile cyber activities and for building long-term cybersecurity strategies. Technological advances are easing this process, as platforms adopt new techniques for the automation of alerts and communications among sharers (Wagner et al., 2019). Thus, in instances when sharing communities are substantively and technologically optimized for cybersecurity, participants benefit from expertise and insights which may otherwise be unavailable to them with respect to developing threat vectors, mitigation of specific cyber risks, and real-time coordinated responses to hostile cyber events.

Nevertheless, together with this chapter's assertion that the use of IS constitutes a best practice and a CBM, IS for the mitigation of cyber risk has also been critiqued for drawbacks and disincentives that have caused the current situation of less than optimal utilization of IS platforms. Some of these challenges – posed to stakeholders that refrain from joining IS platforms, and to IS participants who underuse platforms, or use them as free riders – are reviewed in Section 3. Two of the underlying assumptions of the chapter address this challenge of effective incentivization of stakeholders' use of IS platforms.

The first assumption is that the continued honing of the technological aspects of IS will make platforms more relevant for shareholders: Sharers will increasingly be able to rely upon robust, user-friendly, flexible, and confidential platforms that meet their needs for boosting cybersecurity, especially for coping with real-time cyber events that are in the process of compromising their systems and data. The ongoing relevance and effectiveness of a given IS platform will thus depend upon its incorporation of technology-based best practices for IS, including, *inter alia*, automated threat identification and sharing, vetting of information reliability, and interoperability with other IS platforms.

The second assumption relates to the value of polycentric governance in cyberspace (Craig & Shackelford, 2015). Although no panacea,[3] the sharing of cyber threat information is optimized for platform participants when it engages a plurality and diversity

[3] See below for critique of polycentric governance models in the cybersecurity context in particular; cf. McGinnis (2016).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

of actors: governments, private corporations, NGOs, academia, informal groups, epistemic communities, individuals, and even autonomous or semiautonomous computer systems.[4] Also, optimal IS will include a plurality and diversity of methodologies and measures: real-time information on hostile cyber events, including digital forensics shared by analysts; data on the cyber strategies and policies of private sector organizations, of economic sectors, and of countries; and technical specifications such as those referred to above, evaluations of developing threat vectors, and cyber awareness and training materials. Some of these types of information constitute protected data, the sharing of which impacts substantive legal rights, such as individuals' rights to personal data privacy, corporate intellectual property, and antitrust guarantees (Chabrow, 2015; Elkin-Koren, 1998; Harkins, 2016, pp. 49–63; Shu-yun & Nen-hua, 2007). Analysis of the regulatory protections provided for safeguarding these rights in the context of IS exceeds the scope of the present chapter, and will be treated elsewhere. Support for the position that a polycentric governance model is also advantageous for oversight of such rights protections (Shackelford, 2016) will be expanded upon below.

Thus, to summarize the points raised in this introductory section, we propose in this chapter to show that, to the extent that IS through trusted platforms incorporates modes of polycentric governance, leveraging a multilevel and multisectoral diversity of actors, methodologies, and measures, cybersecurity is supported and the aims of cyber peace are advanced.

In conclusion, an often observed but challenging aspect of cybersecurity and cyber peace in general should also be highlighted in the present IS context: IS is an ongoing exercise in trust building among sharers (Ostrom, Chang, Pennington & Tarko, 1990; Ostrom et al., 2012). Platform participants must be able to rely upon the security of all communications channels, they must have confidence that the data shared will be utilized only in accordance with agreed rules by all participants, and they must have certainty that any stored or retained data are completely protected and that they remain confidential. By leveraging technological developments and modes of polycentric governance, IS has the potential to embody Alexander Klimburg's (2018, p. 359) observation that "trust is a tangible resource in cyberspace," hard coded into its basic protocols, into the development of the Internet and, we venture to add – into secure platforms for the sharing of critical information.

The chapter is structured as follows. Section 2 describes the "how" of IS measures by reviewing selected operational aspects of two examples of IS platforms: one a domestic platform and the second a multilateral one for the global financial sector. Section 3 discusses the ways in which IS mitigates cyber vulnerabilities, and includes some critique of the present utilization of IS. Section 4 characterizes

[4] Such cross-sector cooperation for cybersecurity is becoming increasingly transparent. See, for instance, U.S. Department of Justice (September 16, 2020), and the diversity of participants in the EU's Cyber and Information Domain Coordination Center (https://pesco.europa.eu/project/cyber-and-information-domain-coordination-center-cidcc/).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

the relationship between cyber peace and IS, arguing that IS constitutes a critical building block of sustainable cyber peace governance because of present challenges to binding normative regimes internationally and domestically. Section 5 summarizes the main points and proposes areas for further research that have ramifications for cyber peace IS, including the exploration of IS models with respect to other global collective action problems, such as global health, ensuring global environmental quality, and the elimination of debris in outer space.

2 HOW INFORMATION SHARING WORKS: SELECTED OPERATIONAL ASPECTS OF IS PLATFORMS FOR "BEST PRACTICE" MITIGATION OF CYBER RISK

This section will describe the practical implementation of IS measures by first defining the concept of IS in the cybersecurity context, then noting the key characteristics of IS platforms, before examining two examples of governmental and private sector exchange of cyber information, one domestic in scope (the US' Cyber Information Sharing and Collaboration Program [CISCP]); and the other international and sectoral (Global Financial Services Information Sharing and Analysis Center [FS-ISAC]). The concluding section addresses the operationalization of IS as a standardized best practice for bolstering cybersecurity.

2.1 *Defining Information Sharing*

Information sharing is a measure for interorganizational, intersectoral, and intergovernmental exchange of data that is deemed by sharers to be relevant to the resolution of a collective action problem (Skopik, Settanni, & Fiedler, 2016). In the cyber peace context, it is the agreed upon exchange of an array of cybersecurity related information, such as vulnerabilities, risks, threats, and internal security issues ("tactical IS"), as well as best practices, standards, intelligence, incident response planning, and business continuity arrangements ("strategic IS") (International Standards Organization, 2015). The primary aim of IS in all of these contexts is to reduce information symmetries regarding cyber vulnerabilities at two levels: between hostile cyber actors and their targets and between targeted organizations themselves, none of which has complete situational awareness of the threat environment on their own.[5]

The 2016 *Guide to Cyber Threat Information Sharing*, published by the US National Institute of Standards and Technology (NIST), describes the advantages of IS measures for improving cybersecurity[6] as follows:

[5] Of course, hostile cyber actors also engage in IS, an interesting issue beyond the present scope. See Hausken (2015).

[6] "Cybersecurity" describes the process of applying a "range of actions for the prevention, mitigation, investigation and handling of cyber threats and incidents, and for the reduction of their effects and of the damage caused by them prior, during and after their occurrence." Israeli Government (2015, February 15).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

> By exchanging cyber threat information within a sharing community, organizations can leverage the collective knowledge, experience, and capabilities of that sharing community to gain a more complete understanding of the threats the organization may face. Using this knowledge, an organization can make threat-informed decisions regarding defensive capabilities, threat detection techniques, and mitigation strategies. By correlating and analyzing cyber threat information from multiple sources, an organization can also enrich existing information and make it more actionable (Johnson et al., 2016, p. iii).

These advantages are gained through the resolution of several key issues which arise in defining the four modalities of IS for any given IS platform:

- The agreed rules for thresholds of shared threats and events – IS depends upon the prior agreement among participants as to the threshold events which will trigger the need to share information, especially for the real-time sharing of vulnerabilities and hostile cyber events requiring specific defensive actions such as patching vulnerabilities (ideally within an agreed on window of time). This threshold determination is both substantive and technical: It is set in accordance with legal and regulatory requirements of the given jurisdiction, whether domestic or international, and it is triggered by technical indicators based incident response protocols protecting the network.
- Regulatory issues – Substantive normative and regulatory frameworks constitute an ever-present backdrop for the technological modalities of IS and the determination of IS thresholds. The role of such frameworks in IS, especially the relationship between them and the agreed technical rules for information sharing is critical. They include the aforementioned rights protections (personal data privacy protections, corporate Internet protocol (IP) safeguards, and antitrust guarantees), general international law constraints on hostile cyber activities (Schmitt, 2017), and bilateral and multilateral treaty provisions (Convention on Cybercrime, 2001). Treatment of these substantive issues are beyond the scope of the present chapter and are noted in the Conclusion for further research.
- The types of information shared – Each IS platform specifies the typologies of relevant information to be shared by participants, often in a Terms of Use document that is restricted to the participants – an internal code of conduct that may serve to build trust among sharing entities. Legal and regulatory constraints also determine types of information that may be shared, and the conditions for sharing, such as anonymization of protected personal data. One example is the Cybersecurity Information Sharing Act (2015), S. 754, 114th Cong. (2016), which defines in Section 104(c)(1) two types of shareable information that must be restricted to a "cybersecurity purpose": "cyber threat indicators" and "defensive measures." As discussed below, current developments are moving toward standardization of relevant threat indicators, IS automatization, and rapidity, toward a commoditization of cyber threat data within communities of trust.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

- The sharing entities – Since effective IS platforms are based on communities of trusted sharers, the identity of the sharing entities should be explicit and transparent to all participants (Gill & Thompson, 2016; Lin, Hung, & Chen, 2009; Özalp, Zheng, & Ren, 2014). Moving from the local to the global, sharing of cybersecurity relevant data may take place among individuals (i.e., the MISP and Analyst1 platforms for cyber analysts); within a corporate sector (i.e., the Financial Sector Information and Sharing Analysis Center (FS-ISAC) and Israel's Cyber and Finance Continuity Center (FC3)); between private sector entities and governmental agencies (as in the UK's Cyber Security Information Sharing Partnership [CiSP] and the US' CISCP example below); between one country's governmental agencies (i.e., the US federal government's Cyber Threat Intelligence Integration Center); between states, either bilaterally and multilaterally (i.e., the European Union's CSIRT network as mandated in the Network and Information Systems Directive); and in the framework of international organizations (i.e., NATO's Computer Incident Response Capability).[7]

Moreover, if the definitional scope of IS broadens to include notifications of irregular activity in cyberspace, then sharers also include individual members of the public who may share reports of suspected cyber fraud and cybercrime with entities such as the FBI and national authorities within the EU, via dedicated websites such as the FBI's Internet Crime Complaint Center and the national sites listed on the platform of Europol's Cybercrime Center, "Report Cybercrime Online" (FBI, 2020; Europol, 2020).

The above sampling of sharing entities illustrates the criticality of a polycentric approach to the governance of cyberspace that includes a diversity of actors to address a collective problem. Beyond the modes of IS to bolster cybersecurity among governmental and private companies and organizations reviewed in this Part, current trends in the development of IS include intrasectoral sharing of cyber threat data, integration of artificial intelligence capabilities to improve IS, participation of expert individuals in IS platforms, and the inclusion of the wider public for the purpose of reporting suspicious activity that may constitute a cybercrime, or an indication of a new cyber threat on financial and consumer platforms.

We exclude from the present discussion IS between civilian entities and military or other covert state operators, due to the lack of transparency of most such arrangements (Robinson & Disley, 2010, p. 9). While there are some examples of military actors sharing cyber threat data publicly, as in the US Cyber Command's utilization of the VirusTotal platform in September 2019 to share malware samples associated

[7] There are also open-source sharing communities that make threat indicators publicly available, such as Citizen Lab Reports, (n.d.) and analyst reports that are openly shared online. Such public platforms are definitionally distinct from IS, which relies upon the existence of a closed, trusted community for its effectiveness.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

with the North Korean Lazarus Group, such sharing is neither consistent nor transparent, and thus difficult to analyze conclusively (Vavra, 2019). Should such a trend emerge toward IS by military and intelligence stakeholders with the public, in order to help strengthen common cybersecurity postures, it will be an interesting development that would further support the argument in favor of the polycentricity of IS.

In concluding this initial definitional and conceptual discussion of IS, we note that IS must develop in concert with the changing cyber threat landscape in order to retain its relevance and credibility for participants. These developments dovetail with the approach that cyber peace is a dynamic situation, not a static one, and that it also will take into account changing aspects of cyberspace activities.

In the following two sections, we briefly examine two examples of governmental and private sector exchange of cyber information, each incorporating a different model of IS. The first example, the US' CISCP, constitutes a national platform with both governmental and private sector sharers. The second example, the FS-ISAC, is global in scope[8]; yet, it has been established by private organizations in the financial sector as a not-for-profit entity. Additional platforms, and some of their characteristics, are noted following these two, as well as a brief summary of commonalities and differences.

2.2 *The DHS Cyber Information Sharing and Collaboration Program*

The US Department of Homeland Security and Department of Justice provides a dedicated platform for IS between governmental and private sector organizations, the CISCP. Originally established as a platform for the benefit of critical infrastructure operators pursuant to Presidential Decision Directive-63 of May 1998 (as updated in 2003 by Homeland Security Presidential Decision Directive 7), the CISCP is a generic, voluntary, free-of-charge IS platform, open to public and private sector organizations. By incorporating operators of critical infrastructures and other private and governmental organizations into one platform, CISCP aims "to build cybersecurity resiliency and to harden the defenses of the United States and its strategic partners" (CISCP, www.cisa.gov/ciscp). Thus, it is an explicitly domestic IS platform, operating under US legal and regulatory constraints. Prospective participants sign an agreement establishing the modalities of the exchange of anonymized cybersecurity information, thus ensuring protection from legal liability that may ensue from the sharing of protected information such as personal data, information subject to sunshine laws, and some proprietary data. The platform is described as follows:

> [CISCP] enables actionable, relevant, and timely unclassified information exchange through trusted public-private partnerships across all critical infrastructure … sectors. CISCP fosters this collaboration by leveraging the depth and breadth of DHS cybersecurity capabilities within a focused operational context … [it] helps partners

[8] FS-ISAC headquarters are located in the USA, with offices in the UK and Singapore.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

> manage cybersecurity risks and enhances our collective ability to proactively detect, prevent, mitigate, respond to, and recover from cybersecurity incidents (Cyber Information Sharing and Collaboration Program, www.cisa.gov/ciscp).

Upon completion of an onboarding training session, participating organizations are provided with of two types of CISCP data, reflecting the abovementioned distinction between strategic and tactical IS. The first is ongoing cyber threat information that is made available to participants through indicator bulletins, analysis reports, and malware reports. Two examples are the Weekly Bulletin, summarizing new vulnerabilities according to NIST's National Vulnerability Database classification system (U.S. Department of Homeland Security, 2020) and Joint Alerts, such as that issued in early April 2020 on the exploitation of COVID-19 by malicious cyber actors (Cybersecurity and Infrastructure Agency, 2020b).

The second type of IS provided by CISCP is real-time information about emerging hostile cyber events, characterized by actionable data such as technical indicators of compromise and measures to be taken for resolving them (software updates and patches, file hashes, and forensic timelines). One example is the January 2020 alert regarding serious vulnerabilities in Microsoft Windows operating systems, designated CVE 2020-0601 (also, less officially, "Curveball" and "Chain of Fools") (Wisniewski, 2020). The alert warned of a spoofing vulnerability in the way that Windows validates a certain type of encrypted certificate. A hostile actor could exploit this vulnerability through a man-in-the-middle attack, or by using a phishing website (such as an individual user's bank website) to obtain sensitive financial data or to install malware on a targeted system.

The CISCP shared two types of tactical cybersecurity information with platform participants: A Microsoft Security Advisory addressing the vulnerability by ensuring that the relevant encrypted certificates were completely validated and a National Security Agency advisory providing detection measures for targeted organizations (Cybersecurity and Infrastructure Agency, 2020a). As a result, the Windows vulnerability was quickly identified and addressed by targeted actors. Analysts have noted that IS was especially effective in this incident, resolving a "dangerous zero-day vulnerability" because of the proactive disclosure made by the NSA to Microsoft, and then allowing the vulnerability and patch to be rapidly and simultaneously shared at "machine speed" through the CISCP's automated indicator sharing capability (Wisniewski, 2020). The CVE 2020-0601 event thus exemplifies the importance of leveraging IS among a diversity of sharers – here, governmental and private sector actors – in a transparent manner (Schneier, 2020).

2.3 *Financial Services Information and Analysis Center (FS-ISAC)*

The second IS platform for analysis is FS-ISAC. Like CISCP, it was established pursuant to Presidential Decision Directive-63; yet, the scope of its activity differs from the CISCP in three important respects: It is restricted to the regulated financial

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

sector; it is explicitly global in its membership and scope; and it requires a fee for participation. Thus, it provides a different model for IS from that of the CISCP and focuses on the sector-specific threat vectors and risks of the vulnerable and frequently targeted global financial sector (World Economic Forum, 2019).

FS-ISAC is the leading global IS platform for this sector, which includes 7,000 members in over 70 jurisdictions. It is constituted as a nonprofit organization with headquarters located in the USA and regional hubs in the UK and Singapore. Member institutions are regulated private-sector financial entities (with some exceptions) and include banks, brokerage and securities firms, credit unions, insurance companies, investment firms, payment processors, and financial trade associations. A separate subplatform was established in July 2018 under the auspices of FS-ISAC for governmental and regulatory entities (Cision, 2018): This CERES platform (CEntral banks, REgulators and Supervisory entities) utilizes separate Operating Rules (www.fsisac.com/fsisac-ceres-operating-rules) and Subscriber Agreements (www.fsisac.com/ceres-forum-subscriber-agreement) for its members.

The FS-ISAC platform focuses on intrasectoral IS: The sharing of government sourced information is independently vetted by the platform's Analysis Team as it is shared *via* the DHS' National Cybersecurity and Communications Integration Center, which provides US federal government cyber advisories. The primary objective is to share "relevant and actionable" information among sectoral participants on an ongoing basis "to ensure the continued public confidence in global financial services" (FS-IAC, www.fsisac.com/). The motivation for members to utilize the FS-ISAC platform includes "[its] access to … best-available information, … trusted consultation with other experts in interpreting the information, the classified working environment" (He, Devine, & Zhuang, 2018, p. 217), and the opportunity to access all of this on a single, sector-specific dedicated platform. Shared data include sector-specific threat alerts and indicators, intelligence briefings, tabletop exercises, and mitigation strategies. Participants are eligible to participate in seven separate levels of IS, in accordance with graded membership fee levels, which can amount to tens of thousands of dollars annually (Weiss, 2015, pp. 9–10). To increase its global reach and promote cybersecurity within the financial sector, FS-ISAC also provides a no cost, unidirectional crisis alert service for financial institutions which do not opt for paid membership. The FS-ISAC Operating Rules, Subscriber Terms and Conditions, and End User License Agreement are all available to the public on its website, but those organizations accepted for membership are required to sign an additional, and transparent Subscriber Agreement that is forwarded only following an internal authentication process.

The platform itself is operated by a private sector service provider and overseen by a member constituted board. Information may be attributed or shared anonymously by encrypted web-based connections, and alerts are distributed by the FS-ISAC Analysis Team in accordance with one of the five service levels to which the member has subscribed. Members are notified of urgent and crisis situations via the type

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

of communication they designate (electronic paging, email, Crisis Conference call), and are required by the Subscriber Agreement to access the FS-ISAC portal to retrieve relevant information. Due to the highly regulated nature of the financial sector and the high confidentiality of the information it processes, members are explicitly permitted to submit information anonymously. In addition, all data that have not been specifically designated as attributable to the sharer is subject to a two-step process to scrub all references to the submitting company, one automated via process of keyword search and the second a review by the Analysis Team. Incoming information collected by FS-ISAC from members is shared with government and law enforcement agencies only with consent of the sharing member. Concerns around sharing of sector-specific information are governed by an explicit ban on the exchange of commercial information by antitrust and competition provisions in the Rules and the Subscriber Agreement, and by the applicability of all relevant laws and regulations in member countries (FS-ISAC Operating Rules, art. 9). Likewise, members are bound by a confidentiality agreement and requirements with respect to any sharing of protected personal data (FS-ISAC Operating Rules, arts. 11 & 12).

FS-ISAC maintains an all sector, global cybersecurity alert level, the Financial Services Sector Cyber Threat Advisory, and uses the standardized Traffic Light Protocol (TLP) that is also employed by CISCP, as further described below. Recent research shows that FS-ISAC's use of automated peer-to-peer alerts has decreased the time for generation of cybersecurity compromise indicators by IS participants "from nearly six hours to one minute" (Wendt, 2019a, p. 109), and that "… the automated receipt, enrichment, and triage of [indicators] by the financial institutions were reduced from an average of four hours to three minutes. In total, the automation reduced the average time to produce an IOC, disseminate an IOC, and initiate a response from approximately 10 hours to 4 minutes" (Wendt, 2019b, p. 27).

At present, financial sector entities "actively participate" in peer-to-peer platforms such as FS-ISAC (Wendt, 2019a, p. 115), leveraging automated IS to boost organizational and sectoral cybersecurity. Yet, FS-ISAC and similar sectoral ISACs have come under criticism for the less than optimal participation of members in the platform. Reasons include the platform's reliance on voluntary sharing by members – and thus, the ease with which an institution can act as a "free rider"; the potentially negative impact of sharing of vulnerabilities and risks on commercial reputation and profitability within the sector; and concerns of substantive legal exposures with respect to protected personal data, corporate IP, and antitrust concerns (Liu, Zafar, & Au, 2014, p. 1). The perception of vulnerability given by participation in an IS platform may be an additional factor (Wagner et al., 2019, at 2.6). Thus, on the one hand, the use of FS-ISAC as a platform for sharing among financial sector participants may be readily adopted, especially given the cost-free option made available for receiving urgent governmental alerts. One the other hand, the incentivization of IS on the part of private sector members is much more challenging. We address this concern in Section 4.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

2.4 *Operationalizing IS as a Standardized Best Practice for Cybersecurity*

Information sharing on cyber threats and vulnerabilities of all types that passes through the CISCP, FS-ISAC, and other IS platforms requires technological measures to safeguard IS at three levels: (1) The rapid provision of data by the sharing organization; (2) its confidential transmission; and (3) its timely processing, distribution, and storage on the IS platform. As we have seen in the above examples, IS platforms leverage standardized, automated formats that enable rapid dissemination and reception of cyber threat indicators (CISA Incident Reporting System, www.us-cert.gov/forms/report; US-CERT DHS Cyber Threat Indicator and Defensive Measure Submission System, www.us-cert.gov/forms/share-indicators). Well-known examples are the STIX and TAXII indicator formats[9] that also enable automated information sharing (AIS), Automated Indicator Sharing (AIS), www.us-cert.gov/ais, and the standard TLP, which classifies the security levels of the shared data using four colors in order to indicate the rules for sharing perimeters (see Figure 3.1).[10]

There are many examples of national and transnational IS platforms utilizing similar, standardized systems for threat indicator transmission, including NATO (Oudkerk & Wrona, 2013); the EU's CSIRT network established under the EU NIS Directive (Directive 2016/1148)[11]; the Cyber Threat Alliance (Fortinet, 2017); Israel's "Showcase" (*Chalon Raávah*) (Israel Cyber Directorate, 2019) and its FC3 (Housen-Couriel, 2018; Housen-Couriel, 2019; Ministry of Finance and the Cyber Directorate, 2017); the CiSP of the UK National Cyber Security Center (National Cyber Security Centre, n.d.); and the "Informationspool" platform supported by Germany's Department for Information Sharing (Bundesamt für Sicherheit in der Informationstechnik, BSI) through its "cyber alliance" (Allianz für Cyber-Sicherheit) (Alliance for Cyber Security, n.d.).

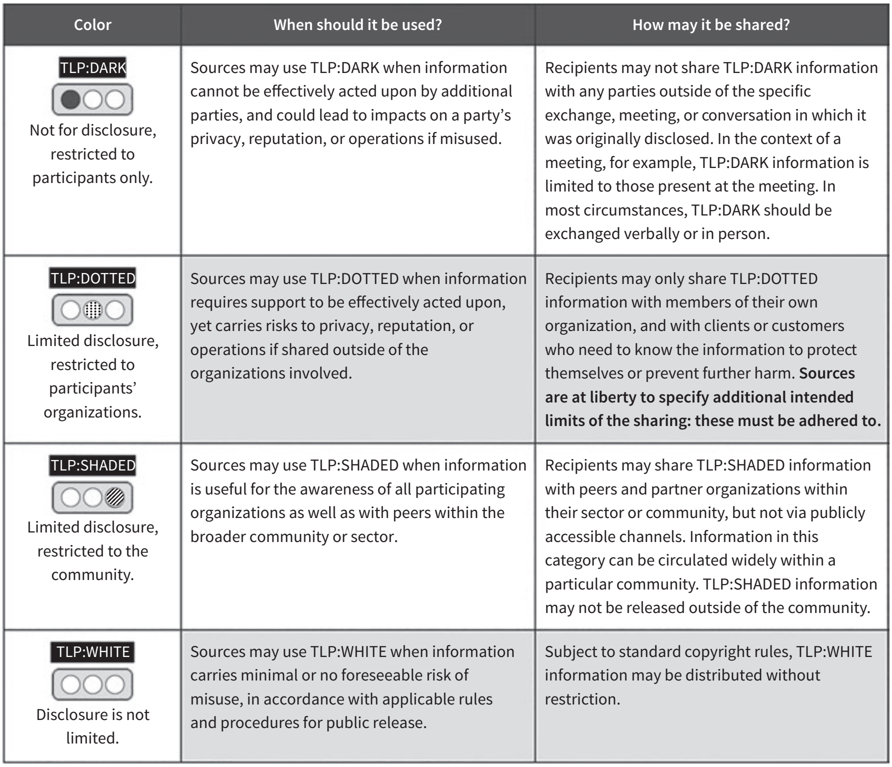
In addition to these IS platforms that foster IS among governmental, corporate, and some other institutional actors for a broad range of cyber threats and risks, several specialized IS platforms focus on a narrower risk typology that pinpoints cybercrime and terrorist activity on the Internet. Examples include INTERPOL's Cybercrime and Cyber-terrorism Fusion Centres (INTERPOL, n.d.); EUROPOL's European Cybercrime Centre (which has been effective in botnet takedown and in the protection of children online) (Europol, n.d.); and the Hash Sharing Consortium established in the framework of the Global Internet Forum to Counter

[9] "STIX is a language … for the specification, capture, characterization and communication of standardized cyber threat information. It does so in a structured fashion to support more effective cyber threat management processes and application of automation." Barnum (2014). See also Van Impe (2015, March 26).

[10] Additional standards are MITRE's Malware Attribute Enumeration and Characterization (MAEC) and OpenIOC, developed by Mandiant (Mavroedis & Bromander, 2017).

[11] The relevant NIS Annex, entitled "Requirements and Tasks of CSIRTs," stipulates their monitoring of risks and incidents; the provision of alerts and other operative indicators; and support for incident response.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D



Color	When should it be used?	How may it be shared?
TLP:DARK ●○○ Not for disclosure, restricted to participants only.	Sources may use TLP:DARK when information cannot be effectively acted upon by additional parties, and could lead to impacts on a party's privacy, reputation, or operations if misused.	Recipients may not share TLP:DARK information with any parties outside of the specific exchange, meeting, or conversation in which it was originally disclosed. In the context of a meeting, for example, TLP:DARK information is limited to those present at the meeting. In most circumstances, TLP:DARK should be exchanged verbally or in person.
TLP:DOTTED ○▦○ Limited disclosure, restricted to participants' organizations.	Sources may use TLP:DOTTED when information requires support to be effectively acted upon, yet carries risks to privacy, reputation, or operations if shared outside of the organizations involved.	Recipients may only share TLP:DOTTED information with members of their own organization, and with clients or customers who need to know the information to protect themselves or prevent further harm. **Sources are at liberty to specify additional intended limits of the sharing: these must be adhered to.**
TLP:SHADED ○○▨ Limited disclosure, restricted to the community.	Sources may use TLP:SHADED when information is useful for the awareness of all participating organizations as well as with peers within the broader community or sector.	Recipients may share TLP:SHADED information with peers and partner organizations within their sector or community, but not via publicly accessible channels. Information in this category can be circulated widely within a particular community. TLP:SHADED information may not be released outside of the community.
TLP:WHITE ○○○ Disclosure is not limited.	Sources may use TLP:WHITE when information carries minimal or no foreseeable risk of misuse, in accordance with applicable rules and procedures for public release.	Subject to standard copyright rules, TLP:WHITE information may be distributed without restriction.

FIGURE 3.1 Traffic Light Protocol (TLP) definitions and usage, CISA [no date].

Terrorism (GIFCT) founded in 2016 by Facebook, Google, YouTube, and Twitter to share information on extremist and terrorist content online and containing more than 200,000 such hashes (Global Internet Forum to Counter Terrorism, n.d.).

These and other such IS platforms reflect organizational and regional differences in the modes of gathering and processing cyber threat indicators and other operational data. Yet, they all rely on standardized and vetted processes that promote trust among sharing entities (International Standards Organization, 2015). The developing technical protocols and the informal codes of conduct around their use constitute an important aspect of IS as a best practice for cybersecurity, and contribute to incentivizing it for use by a plurality of sharers.

3 MITIGATION OF CYBER THREATS AND EVENTS THROUGH INFORMATION SHARING: DISCUSSION

Although neither the sole means of closing gaps in cybersecurity, nor by any means a blanket remedy, IS already serves as a key measure for bolstering national, sectoral and, ultimately, global cybersecurity by leveraging and optimizing

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

interdependencies (Europol, 2017). Nevertheless, there is still critique of its present use as a measure for boosting cybersecurity and mitigating risk.[12] Melissa Hathaway (2010) has noted that the considerable quantity of available IS platforms poses a challenge for limited organizational and governmental resources, causing confusion and under commitment (counting fifty-five such government initiated partnerships in the USA alone). Zheng and Lewis (2015, p. 2) emphasize "programmatic, technical and legal challenges" to IS. Lubin (2019) posits that the increased adoption of cyber insurance policies by private corporations, groups, and individuals may have a chilling effect on IS because "there are often very strict parameters regarding [a policy holder's] notification and cooperation [regarding hostile cyber events] in the insurance policy." Finally, the methodologies for evaluating the success of certain IS platforms over others are still developing – as are the definitions of "success" itself in the cyber context (Garrido-Pelaz, González-Manzano, & Pastrana, 2016, pp. 15–24).

The reasons that organizations may fail to fully adopt and operationalize IS, despite its advantages, may be characterized as either (1) operative or (2) normative-substantive.

The operative disincentives include:

- The inability to establish trust among sharing entities, some of whom may be competitors, including the concern regarding free riders (entities who benefit from IS without contributing themselves).
- Costs related to IS including recruitment, training and retention of appropriate cybersecurity personnel and organizational time spent on IS, including time devoted to "false positives" (i.e., incorrect alerts that are based on bad information) (Powell, 2005, p. 507).
- Lack of transparency regarding the robustness and confidentiality of IS platforms, including the possible use of shared data by any participating government agencies for noncybersecurity purposes, such as the tracking of individuals for immigration control or unauthorized surveillance (Johnson et al., 2016, pp. 4–5).
- Regulatory redundancy, where other, possibly competing, IS formats are mandated and may complicate efficient IS (Knerr, 2017, pp. 550, 553; Robinson, 2012).[13]
- Concern that participation in IS platforms may result in the perception that the sharer is vulnerable to cyber threats (Wagner et al., 2019, at 2.6).

[12] The well-known example of the 2017 breach into the Equifax credit reporting company illustrates the pitfalls that characterize the reluctance of some financial sector actors to engage effectively with IS. See Warren (2018). See also Fournoy & Sulmeyer (September/October, 2018).

[13] One leading example can be seen in the USA, where the financial sector is defined as one of the sixteen included under the aegis of DHS and also subject to the directives of the US Department of Treasury and anti-money laundering reporting requirements.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Three of the normative-substantive disincentives are:

- The potential exposure of protected personal data shared by organizations, with resulting regulatory sanctions and exposure to litigation by data subjects and regulators.
- The potential exposure of organizational IP, with potential chilling effects on organizational innovation, and possible implications for corporate market value.
- Concerns regarding antitrust implications of IS within a sector.

Taken together, both the operative and substantive-normative disincentives to IS help to explain why some cyberspace actors are reluctant to fully adopt IS as part of their overall cybersecurity strategies on their own initiative; and when they participate, may do so less than optimally (including in situations where required to do so by regulators) (Barford et al., 2010, pp. 3–13; Sutton, 2015, pp. 113–116). Nonetheless, despite these potential weaknesses in IS platforms, there is, overall, strong continued support for their inclusion in legal, policy, and standardization initiatives, as shall be shown in the following section. Not only do the potential advantages of increased "cyber situational awareness" outweigh the disincentives but, as argued here, technological developments such as standardized reporting of cyber threat indicators, STIX and TAXII architectures, TLP, and increasingly automated IS (the "commoditization" of cyber threat indicators) signal an increasing awareness of the criticality of IS for the mitigation of cyber risk on the part of all stakeholders.

4 CHARACTERIZING THE RELATIONSHIP BETWEEN CYBER PEACE AND INFORMATION SHARING: A BEST PRACTICE AND CONFIDENCE-BUILDING MEASURE THAT LEVERAGES POLYCENTRICITY

4.1 *Information Sharing as a Best Practice in Support of Cyber Peace*

The definition of cyber peace cited at the beginning of this chapter identifies four of its aspects: clarification of "rules of the road" for setting actors' expectations and thresholds for IS; threat reduction; risk assessment; and best practices for carrying out these three tasks – all of which are supported by IS. Participants in any given IS platform agree *ex ante* to the *thresholds of nonpermissible online behavior* of hostile actors, by virtue of the triggers indicating precisely when relevant information should be shared by them and is shared with them. Typical *informational asymmetries* that have characterized cyber hostilities to the advantage of the attacker are addressed by the sharing of data, such as by those alerts referred to in the above examples of CISCP and FS-ISAC. *Risk assessment* is carried out, *inter alia*, on the basis of indicators, data, and situational evaluations received through IS.

Two additional attributes of IS that support sustainable and scalable cyber peace should be noted. First, its neutrality with respect to the typology of both attackers

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

and targets. Whether the attacker is an individual, a country, a group of criminal hackers, an inside operator, or an autonomous or semiautonomous computer – the IS alert thresholds are similar.[14] Likewise, alerts, vulnerabilities, and warnings are target neutral, and are similarly applicable in the context of state-to-state hostilities, cybercrime, terrorist activity, hacktivism, and money laundering. The second attribute is the convenient scalability of IS, as sharing technologies and protocols currently undergo standardization, automatization, and commoditization.

Work is still needed to quantify the specific advantages that IS brings as a best practice in boosting levels of cybersecurity, especially in terms of its cost effectiveness as part of the overall cybersecurity strategy of organizations and states. This much needed analysis will contribute to a better understanding of the economic aspects of sustainable cyber peace, as well.

4.2 *Beyond Best Practice: The Value of Information Sharing as a CBM*

Building on this understanding of IS as a best practice, it is argued here that IS further supports sustainable cyber peace as a CBM at the international level, among the states, international organizations, and multinational companies that are critical to ensuring global cybersecurity. The framing of IS as a CBM, rather than as a binding, substantive norm to which these entities are subject as a matter of law or policy, is beneficial to the utilization of IS platforms at the international level (Borghard & Lonergan, 2018). By sidestepping substantive multilateral commitments, IS can be more readily utilized to support cybersecurity and cyber peace. Examples where this has occurred include the UN's 2015 GGE (United Nations General Assembly, 2015), the OSCE's 2016 listing of cybersecurity CBMs (Organization for Security and Co-Operation in Europe, 2016), and the 2018 Paris Call for Trust and Security in Cyberspace (Principle 9).

CBMs were originally used in the context of the Cold War to further disarmament processes in the context of the diplomatic and political standoff between the USSR and the West. Nonmilitary CBMs have been defined more generally as "actions or processes undertaken … with the aim of increasing transparency and the level of trust" between parties (Organization for Security and Co-operation in Europe, 2013). They are "one of the key measures in the international community's toolbox aiming at preventing or reducing the risk of a conflict by eliminating the causes of mistrust, misunderstanding and miscalculation" (Pawlak, 2016, p. 133). CBMs are also critical in the global cybersecurity context and have been described as a "key tool in the cyber peacebuilder's toolkit" (Nicholas, 2017).

In a 2017 in-depth study of eighty-four multilateral and bilateral initiatives addressing the collective action challenges of cybersecurity, including treaties,

[14] Barring, of course, attacks which protected systems have been directed to ignore such as pentesting and friendly intrusions. These are not always transparent to IS participants.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

codes of conduct, agreements, memoranda and public declarations, IS was found to be included as an agreed cybersecurity measure in more than 25 percent of such initiatives (twenty-one out of the total eighty-four) (Housen-Couriel, 2017, pp. 51–52). Moreover, the analysis was able to isolate several specific elements of IS, discussed above, that were individually included in this top quarter: IS measures in general[15]; establishment of a specific national or organizational point of contact for information exchange; and sharing of threat indicators (Housen-Couriel, 2017, pp. 51–52).[16] These elements were three out of a list of a dozen CBMs that occur with sufficient frequency to be included in a "convergence of concept" with which diverse stakeholders – states, regional organizations, intergovernmental organizations, specialized UN agencies, standards organizations, private corporations, sectoral organizations, and NGOs – have incorporated into cybersecurity initiatives.[17] The study concluded that, while such cyberspace stakeholders are frequently willing to incorporate general arrangements for IS (it is in fact the leading agreed-upon cyber CBM in the initiatives that were studied), and even to specify a national or organizational point of contact, they are less willing to commit to a 24/7, real-time exchange of cybersecurity related information (Housen-Couriel, 2017, p. 67). This finding indicates a gap that should be considered in the context of further leveraging IS in the context of cyber peace.

Nonetheless, as noted above, IS as a CBM holds the advantage of bypassing the present, considerable challenges of achieving formal and substantive multistakeholder agreement on substantive cyber norms, until such time as such binding norms are legally and geopolitically practicable (Efroni & Shany, 2018; Finnemore & Hollis, 2016; Macak, 2017). A few examples of binding domestic law and international regulatory requirements for organizational participation in IS platforms do exist, such as the pan-EU regime established under the EU NIS (Directive 2016/1148), the Estonian Cybersecurity Act of 2016, and the US Department of Defense disclosure obligations for contractors when their networks have been breached. However, there are many more based on voluntary participation, such as the CISCP and FS-ISAC reviewed above, Israel's FC3, and the global CERT and CSIRT networks of 24/7 platforms for cyber threat monitoring, including the EU network of more than 414 such platforms (European Union Network and Information Security Agency, 2018).

[15] Defined as "exchange between stakeholders of information about strategies, policies, legislation, best practices, and cyber infrastructure capacity building." Forty-three out of the eighty-four included this measure.

[16] Twenty-three out of the eighty-seven included this measure, and eighteen out of eighty-four included real-time 24/7 exchange of threat data.

[17] These are: Information sharing, in general, sharing of information around cyber threats, law enforcement cooperation, protection of critical infrastructure, mechanisms for cooperation with the private sector and civil society, arrangements for international cooperation, a mechanism for vulnerability disclosure, regular dialogue, the mandating of general legislative measures, training of cyber personnel, cyber education programs, and conducting tabletop exercises.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

For the purposes of its analysis in this chapter, IS constitutes as a nonbinding CBM that also constitutes a best practice for bolstering cybersecurity and cyber peace, yet does not require a binding legal basis for its implementation. The critical issue of the use of regulatory measures, both binding and voluntary, to promote IS for optimal cybersecurity and cyber peace is, as noted above, an issue for further research.

4.3 *Leveraging Polycentricity for Effective IS*

In this section, we briefly address the advantages of a polycentric approach for effective IS. Polycentricity is an approach and framework for ordering the actions of a multiplicity and diversity of actors around a collective action problem.[18] Several scholars in the field of cybersecurity describe and analyze regulatory activity in cyberspace specifically in accordance with such an approach (Craig & Shackelford, 2015; Kikuchi & Okubo, 2020; Shackelford, 2014, pp. 88–108). Polycentricity explicitly recognizes a multiplicity of sources of regulatory authority and behavioral organization for cyber activities, including nation-state actors, private sector organizations, third sector entities, and even individuals, and it acknowledges the value of employing a diversity of measures to address the collective action problem (Elkin-Koren, 1998; Shackelford, 2014; Thiel et al., 2019).

A polycentric approach is theoretically and conceptually most appropriate for supporting IS in particular and cybersecurity overall due, *inter alia*, to its inherent stakeholder inclusiveness, flexibility with regard to types of regulatory measures, and transparency with respect to potential violations of substantive privacy rights, IP protections, and antitrust provisions (Shackelford, 2014, p. 107). Moreover, in the context of IS, a polycentric approach maximizes the potential for remedying informational asymmetries among a diversity of vetted sharers, bringing to bear a variety of perspectives and capabilities (Kikuchi & Okubu, 2020, pp. 392–393; Shackelford, 2013, pp. 1351–1352).[19] Such an approach explicitly acknowledges the complex interdependencies of all actors in cyberspace (Shackelford, 2014, pp. 99–100). Thus, a polycentric approach will optimally include on an IS platform the broadest possible

[18] Polycentricity is "a system of governance in which authorities from overlapping jurisdictions (or centers of authority) interact to determine the conditions under which these authorities, as well as the citizens subject to these jurisdictional units, are authorized to act as well as the constraints put upon their activities for public purposes." McGinnis (2011), pp. 171–72. See also Black (2008), p. 139 ("'Polycentric regulation' is a term which acts … to draw attention to the multiple sites in which regulation occurs at sub-national, national and transnational levels.")

[19] Specifically, key parameters include the explicit inclusion of a multiplicity and diversity of trusted participants, and a range of regulatory incentives, tools and measures employed for IS. These might encompass, *inter alia*, national laws, sectoral self-regulation, best practices, guidelines, standards, international agreements, public–private partnerships, academic and consulting reports, and other types of regulation through information sharing. On the other hand, some drawbacks to the polycentric approach include fragmentation, "gridlock," inconsistency, and "the difficult task of getting diverse stakeholders to work well together across sectors and borders."

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

range of sharers: Government regulators and agencies themselves; sectoral actors that may share information informally, as they are targeted simultaneously by malicious cyber actors; umbrella groups formed within the sector for formal and informal IS; technical experts, academic and consulting actors, providing external assessments of IS models and their effectiveness; and individuals who may share information through governmental, sectoral, or organizational channels, or through informal channels such as social media – when they experience compromised cybersecurity through their personal Internet use.

The two examples reviewed above are relatively non polycentric at present: CISCP is a public–private sector partnership that includes government agencies and companies in its membership, and FS-ISAC restricts participation even further, to private sector members only (central banks, sector regulators, and other government agencies must join the separate CERES platform). The challenges for building trust on these two platforms are significant and may continue to constitute barriers for inclusion of a broader, more diverse membership. In the context of the financial sector, especially, a more polycentric participation in IS may be encumbered at present by legal and regulatory constraints. Nevertheless, financial institutions already recognize the important potential of gathering data on unusual, detrimental activity in their networks *via* reporting by customers and suppliers – that is, individual users who access parts of the network regularly and often, and who can serve as sensors for fraudulent and hostile cyber activity such as phishing (Cyber Security Intelligence, 2017). Individual user endpoints and accounts may be among the most vulnerable points of entry into an institution's network, but they also constitute a key element for cybersecurity data gathering at the perimeter of financial institutions that, we contend, should be leveraged within IS platforms as an additional means of mitigating the informational asymmetry between the hostile actor and the targeted organization. Thus, the provision of fraud prevention alert mechanisms on the websites of banks and some other private companies, by means of which customers may provide information about phishing schemes, irregular activity in their accounts, and other suspicious activity, might be incorporated into sectoral IS platforms.[20] This growing understanding on the part of financial organizations, social media platforms, and consumer websites that much valuable information with respect to cyber risks may be garnered from individuals (including customers, employees, and suppliers) requires creative thinking around the incentivization of such IS, as well as the protection of individual privacy rights as cyber risk indicators are shared.[21]

[20] See, for example, the portals for reporting suspicious cyber activity at amazon.com (www.amazon.com/gp/help/customer/display.html/ref=hp_left_v4_sib?ie=UTF8&nodeId=GPXKBLY3LY4ZNG5H); Bank of America (www.bankofamerica.com/security-center/report-suspicious-communications/#:~:text=Forward%20any%20suspicious%20email%20or,at%20800%2D432%2D1000.); and the Internal Revenue Service (www.irs.gov/privacy-disclosure/report-phishing).

[21] A key challenge in this context is the evolution of full, mutual IS, and not only unilateral reporting of risks on the part of individuals to their banks, social media platforms, and consumer platforms.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

In summary, IS is likely be most effective as best practice at the domestic level and as a CBM at the international level – when it is governed by a polycentric approach for the most efficient pooling of resources, knowledge, and experience to mitigate, counter, and respond effectively to cyber threats and events.

5 SUMMARY AND CONCLUSIONS

This chapter has aimed to show how IS platforms can serve as: arbiters of cyber expertise; the exchange of technical data; real-time coordination of defensive actions; and, perhaps most importantly, the development of trust among key stakeholders in order to mitigate the effects of hostile activities in cyberspace. The analysis has aimed to support the thesis that one of the critical elements to achieving sustainable cyber peace, indeed a *sine qua non* for its governance, is the timely utilization of credible IS platforms that allow entities targeted by hostile cyber activities to pool information, resources, and insights in order to mitigate cyber risk. Successful platforms will leverage innovative technological developments for collecting actionable cyber threat data at both the tactical, real-time level of incident response, as well as that of strategic planning for amending vulnerabilities and developing long-term defense strategies.

Moreover, even as IS modalities are included in many initiatives for promoting cybersecurity among state and nonstate actors, they have the advantage of bypassing need to achieve formal and substantive multistakeholder agreement on cyber norms that are at the core of international and domestic legal regimes for the governance of cyberspace. At the international level, many contemporary scholars note that the difficulties of surmounting normative barriers await resolution until such time as states and international organizations are prepared to act more transparently in cyberspace and forge binding international and domestic legal regimes. Eventually, in international regimes to which states and organizations formally agree – or, perhaps, more gradually through the evolution of international custom – IS may be transformed from a norm-neutral CBM into an element of states' and organizations' due diligence under international cyber law.[22]

Several issues that are beyond the present scope of this chapter invite additional research. Among them are the quantifiable, cost–benefit calculations of IS platforms as an element of cybersecurity and cyber peace; the role of regulation (including substantive legal norms) in promoting and incentivizing IS; the cumulative effects of standardization and automatization on IS processes; and a broader examination of the specific advantages of an explicitly polycentric approach to IS. IS models with respect to other global collective action problems, such as public health (especially relevant in the present COVID-19 pandemic), environmental quality, and the elimination of

[22] On aspects of due diligence in the context of international cyber law, see Tallinn 2.0, Rules 6 and 7 at 30–50, and Rule 6 at 288.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

outer space debris are also salient: A broader, comparative analysis of IS regimes for the mitigation of risk in meeting these common problems may prove fruitful.

We conclude with a note of deep appreciation for the talented and committed women and men who are the ultimate heroes of the story of cyber IS: The security analysts who mine, winnow, and share critical cyber threat indicators as a matter of course, 24 hours a day, 365 days a year, over weekends, during their holiday breaks, and from anywhere they can possibly connect up to cyberspace.

REFERENCES

Ablon, L., & Bogart, A. (2017). *Zero days, thousands of nights: The life and times of zero-Day vulnerabilities and their exploits*. RAND Corporation. www.rand.org/pubs/research_reports/RR1751.html

Alliance for Cyber Security. (n.d.). *Informationspool*. Retrieved October 24, 2020 from www.allianz-fuer-cybersicherheit.de/ACS/DE/Informationspool/_function/Informationspool_Formular.html;jsessionid=44A7CF329463873BACD747ABEBA5CB17.1_cid351?nn=6643342

Barford, P., Dacier, M., Dietterich, T., Fredrikson, M., Giffin, J., Jajodia, S. et al. (2010). Cyber SA: Situational awareness for cyber defense. In S. Jajodia, P. Liu, V. Swarup, & C. Wang (Eds.), *Cyber situational awareness, advances in information security* (pp. 3–13).

Barlow, J. P. (1996). *Declaration of the independence of cyberspace*.

Barnum, B. (2014). *Standardizing cyber threat intelligence information with the structured threat information eXpression (STIX™)*. http://stixproject.github.io/about/STIX_Whitepaper_v1.1.pdf

Black, J. (2008). Constructing and contesting legitimacy and accountability in polycentric regulatory regimes. *Regulation & Governance*, *2*(2), 137–164.

Borghard, E., & Lonergan, S. (2018). Confidence building measures for the cyber domain. *Strategic Studies Quarterly*, *12*(3), 10–49. www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-12_Issue-3/Borghard-Lonergan.pdf?ver=fvEYs48lWSdmgIJlcAxPkA%3d%3d

Chabrow, E. (2015, March 15). *Cyberthreat information sharing privacy concerns raised*. BankInfoSecurity. www.bankinfosecurity.com/privacy-risks-raised-over-cyberthreat-information-sharing-a-8970

Cision. (2018, June 11). *FS-ISAC launches the CERES forum: World's First Threat Information Sharing Group for Central Banks, Regulators and Supervisors*. www.prnewswire.com/news-releases/fs-isac-launches-the-ceres-forum-worlds-first-threat-information-sharing-group-for-central-banks-regulators-and-supervisors-300663921.html

Citizen Lab Reports. (n.d.). *Targeted threats*. Retrieved October 24, 2020 from https://citizenlab.ca/category/research/targeted-threats/

Convention on Cybercrime. (2001, November 23*)*. E.T.S. No. 185.

Craig, A., & Shackelford, S. (2015). Hacking the planet, the Dalai Lama, and You: Managing technical vulnerabilities in the internet through polycentric governance. *Fordham Intellectual Property, Media & Entertainment Law Journal*, *24*(2), 381–425.

Cyber Security Intelligence. (2017, May 1). *The cyber security threats that keep banks alert*. www.cybersecurityintelligence.com/blog/the-cybersecurity-threats-that-keep-banks-alert-2392.html

Cybersecurity Act of 2018. (2018, May 23). www.riigiteataja.ee/en/eli/523052018003/consolide

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Cybersecurity and Infrastructure Agency. (2020a, April 8). *Alert (AA20-009A): Covid-19 exploited by malicious cyber actors*. www.us-cert.gov/ncas/alerts/aa20-099a

Cybersecurity and Infrastructure Agency. (2020b, January 14). *Alert (AA20-014A): Critical vulnerabilities in microsoft windows operating system*. www.us-cert.gov/ncas/alerts/aa20-014a

Deljoo, A., van Engers, T., Koning, R., Gommans,L., & de Laat, C. (2018). *Towards trustworthy information sharing by creating cyber security alliances*. 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE), 1506–1510.

Directive 2016/1148, of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the union, 2016 O.J. (L194) 1.

Efroni, D., & Shany, Y. (2018). A rule book on the shelf? Tallinn Manual 2.0 on cyber operations and subsequent state practice. *American Journal of International Law, 112*(4), 583–657.

Elkin-Koren, N. (1998). Copyrights in cyberspace – Rights without laws. *Chicago-Kent Law Review*, 73(4), 1156–1201.

European Union Network and Information Security Agency. (2018). *Cooperative models for Information Sharing and Analysis Centers (ISACs)*. www.enisa.europa.eu/publications/information-sharing-and-analysis-center-isacs-cooperative-models

Europol. (2017, December 4). *Andromeda botnet dismantled in international cyber operation*. [Press release]. www.europol.europa.eu/newsroom/news/andromeda-botnet-dismantled-in-international-cyber-operation

Europol. (n.d.). *EC3-European cyber crime centre*. www.europol.europa.eu/about-europol/european-cybercrime-centre-ec3

Europol. (2020). *Report Cybercrime Online*. www.europol.europa.eu/report-a-crime/report-cybercrime-online

FBI. (2020, May 8). *The FBI's Internet Crime Complaint Center (IC3) marks its 20th Year* [Press release]. www.fbi.gov/news/pressrel/press-releases/the-fbis-internet-crime-complaint-center-ic3-marks-its-20th-year

Finnemore, M., & Hollis, D. (2016). Constructing norms for global cybersecurity. *American Journal of International Law*, 110(3), 425–479.

Fortinet. (2017, February 14). *Cyber threat alliance expands mission through appointment of President, formal incorporation as not-for-profit and new founding members* [Press release]. www.fortinet.com/ru/corporate/about-us/newsroom/press-releases/2017/cyber-threat-alliance-expands-mission.html

Fournoy, M., & Sulmeyer, M. (2018, September/October). Battlefield internet: A plan to secure cyberspace. *Foreign Affairs*. www.foreignaffairs.com/articles/world/2018-08-14/battlefield-internet.

Garrido-Pelaz, R., González-Manzano, L., & Pastrana, S. (2016). *Shall we collaborate? A model to analyse the benefits of information sharing* [Workshop presentation]. Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security.

Gill, R., & Thompson, M. (2016). *Trust and information sharing in multinational-multiagency teams*. Springer.

Global Internet Forum to Counter Terrorism. (n.d.). *Joint tech innovation*. Retrieved October 24, 2020 from https://gifct.org/joint-tech-innovation/

Harkins, M. W. (2016). *Managing risk and information security*. Apress.

Hathaway, M. (2010, May 7). *Why successful partnerships are critical for promoting cybersecurity*. Executive Biz.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Hausken, K. (2015). A strategic analysis of information sharing among cyber hackers. *Journal of Information Systems and Technology Management*, *12*.

He, M., Devine, L., & Zhuang, J. (2018). Perspectives on cybersecurity information sharing among multiple stakeholders using a decision-theoretic approach. *Risk Analysis*, 38(2), 215–225.

Housen-Couriel, D. (2017). *An analytical review of and comparison of operative measures included in cyber diplomatic initiatives* (GCSC Issue Brief No. 1). Global Commission on the Security of Cyberspace.

Housen-Couriel, D. (2018). Information sharing for mitigation of hostile activity in cyberspace (Part 1). *European Cybersecurity Journal*, 4(3), 44–50.

Housen-Couriel, D. (2019). Information sharing for mitigation of hostile activity in cyberspace (Part 2). *European Cybersecurity Journal*, 5(1), 16–24.

International Standards Organization. (2015). *ISO/IEC 27010:2015, Information Technology – Security Techniques – Information security management for inter-sector and inter-organizational communications*. www.iso.org/standard/44375.html

INTERPOL. (n.d.). *Cybercrime*. www.interpol.int/content/download/5267/file/Cybercrime.pdf

Israel Cyber Directorate. (2019). *Israel's 'Showcase' for evaluation of cyber risks*. www.gov.il/he/departments/general/systemfororg

Israeli Government. (2015, February 15). *Resolution No. 2444, advancing the national preparedness for cyber security*.

Johnson, C., Badger, L., Waltermire, D., Snyder, J., & Skorupka, C. (2016). *Guide to cyber information threat sharing* (NIST Special Pub. 800-150). National Institute of Standards & Technology. http://dx.doi.org/10.6028/NIST.SP.800-150

Kikuchi, M., & Okubo, T. (2020). Building cybersecurity through polycentric governance. *Journal of Communications*, 15, 390–397.

Klimburg, A. (2018). *The darkening web: The war for cyberspace*. Penguin Books.

Knerr, M. (2017). Password please: The effectiveness of New York's first-in-nation cybersecurity regulation of banks. *Business Entrepreneurship & Tax Law Review*, *1*(2), 539–555.

Lin, M. J. J., Hung, S. W., & Chen, C.J. (2009). Fostering the determinants of knowledge sharing in professional virtual communities. *Computers in Human Behavior*, 25(4), 929–939.

Liu, C. Z., Zafar, H., & Au, Y. (2014). Rethinking FS-ISAC: An IT security information sharing network model for the financial services sector. *Communications of the Association for Information Systems*, 34(1).

Lubin, A. (2019, September 21). *The insurability of cyber risk* [Unpublished manuscript]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3452833

Macak, K. (2016). Is the international law of cyber security in crisis? In N. Pissanidis, H. Rõigas, & M. Veenendaal (Eds.), *Cyber power* (pp. 127–140). NATO CCD COE Publications.

Macak, K. (2017). From cyber norms to cyber rules: Re-engaging states as law-makers. *Leiden Journal International Law*, 30(4), 877–899.

Mavroedis, V., & Bromander, S. (2017). *Cyber threat intelligence model: An evaluation of taxonomies, sharing standards and ontologies within cyber threat intelligence*. IEEE 2017 European Intelligence and Security Informatics Conference, 91–98.

McGinnis, M. (2011). An introduction to IAD and the language of the Ostrom Workshop: A simple guide to a complex framework. *Policy Studies Journal*, 39(1), 169–183.

McGinnis, M. (2016). *Polycentric governance in theory and practice: Dimensions of aspiration and practical limitations*. https://mcginnis.pages.iu.edu/polycentric%20governance%20theory%20and%20practice%20Feb%202016.pdf

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Ministry of Finance and the Cyber Directorate. (2017, September 4). *Memorandum from the finance cyber and continuity centre (FC3)*. https://docs.google.com/viewer?url=http%3A%2F%2Fwww.export.gov.il%2Ffiles%2Fcyber%2FFC3.PDF%3Fredirect%3Dno

National Cyber Security Centre. (n.d.). *CiSP terms and conditions* (v.5). www.ncsc.gov.uk/files/UK%20CISP%20Terms%20and%20Conditions%20v5.0.pdf

Nicholas, P. (2017, June 29). *What are confidence building measures (CBMs) and how can they improve cybersecurity?* Microsoft. www.microsoft.com/en-us/cybersecurity/blog-hub/CMB-and-cybersecurity

Organization for Security and Co-Operation in Europe. (2013). *OSCE guide on non-military confidence-building measures (CBMs)*. www.osce.org/secretariat/91082

Organization for Security and Co-Operation in Europe. (2016, March). *Decision No. 1202, confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies*. https://ccdcoe.org/incyder-articles/osce-expands-its-list-of-confidence-building-measures-for-cyberspace-common-ground-on-critical-infrastructure-protection/

Ostrom, E., Chang, C., Pennington, M., & Tarko, V. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E., Chang, C., Pennington, M., & Tarko, V. (2012). *The future of the commons: Beyond market failure and government regulation*. The Institute of Economic Affairs.

Oudkerk, S., & Wrona, K. (2013). Using NATO labelling to support controlled information sharing between partners. In E. Luiijf & P. Hartel (Eds.), *Critical information infrastructures security, lecture notes in computer science* (Vol. 8328). Springer Link.

Özalp, Ö., Zheng, Y., & Ren, Y. (2014). Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. *Management Science*, 60(10), 2435–2460. https://doi.org/10.1287/mnsc.2014.1905

Paris Call for Trust and Security in Cyberspace. (2018, November 12). https://pariscall.international/en/

Pawlak, P. (2016). Confidence building measures in cyberspace: Current debates and trends. In A.-M. Osula & H. Rõigas (Eds.), *International cyber norms: Legal, policy & industry perspectives* (pp. 129–153). CCDCOE.

Powell, B. (2005). Is cybersecurity a public good? Evidence from the financial services industry. *Journal of Law, Economics & Policy*, 1(2), 497–510.

Presidential Decision Directive PDD/NSC 63. (1998, May 22). https://fas.org/irp/offdocs/pdd/pdd-63.htm

Robinson, N. (2012). Information sharing for CIP: Between policy, theory, and practice. In C. Laing, A. Baadi, & P. Vickers (Eds.), *Securing critical infrastructures and critical control systems: Approaches for threat protection*. IGI Global.

Robinson, N., & Disley, E. (2010). *Incentives and challenges for information sharing in the context of network and information security*. European Union Network and Information Security Agency. www.enisa.europa.eu/publications/incentives-and-barriers-to-information-sharing

Ruhl, C., Hollis, D., Hoffman, W., & Maurer, T. (2020). *Cyberspace and geopolitics: Assessing global cybersecurity norm processes at a crossroads*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2020/02/26/cyberspace-and-geopolitics-assessing-global-cybersecurity-norm-processes-at-crossroads-pub-81110

Schmitt, M. (Ed.). (2017). *Tallinn Manual 2.0 on the international law applicable to cyber operations* (2nd ed.). Cambridge University Press.

Schneier, B. (2020, January 15). *Critical windows vulnerability discovered by NSA*. Schneier on Security. www.schneier.com/blog/archives/2020/01/critical_window.html

Shackelford, S. (2013). Toward cyberpeace: Managing cyberattacks through polycentric governance. *American University Law Review*, 62(5), 1273–1364.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Shackelford, S. (2014). *Managing cyber attacks in international law, business, and relations: In search of cyber peace*. Cambridge University Press.

Shackelford, S. (2016). Protecting intellectual property and privacy in the digital age: The use of national cybersecurity strategies to mitigate cyber risk. *Chapman Law Review*, 19(2), 445–482.

Shu-yun, Z., & Neng-hua, C. (2007). *The collision and balance of information sharing and intellectual property protection*. http://en.cnki.com.cn/Article_en/CJFDTOTAL-TSGL200702010.htm

Skopik, F., Settanni, G., & Fiedler, R. (2016). A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security*, 60, 154–176.

Sutton, D. (2015). Trusted information sharing for cyber situational awareness. *E & I Elektrotechnik und Informationstechnik*, 132(2), 113–116.

Thiel, A., Garrick, D., & Blomquist, W. (Eds.). (2019). *Governing complexity: Analyzing and applying polycentricity*. Cambridge University Press.

United Nations General Assembly. (2015, July 22). Report A/70/174: Report of the group of governmental experts on developments in the field of information and telecommunications in the context of international security. http://undocs.org/A/70/174

U.S. Department of Homeland Security. (2020, March 30). Bulletin SB-20-097. www.us-cert.gov/ncas/bulletins/sb20-097

U.S. Department of Justice. (2020, September 16). *Seven international cyber defendants, including "Apt 41" actors, charged in connection with computer intrusion campaigns against more than 100 victims globally* [Press Release]. www.justice.gov/opa/pr/seven-international-cyber-defendants-including-apt41-actors-charged-connection-computer

Van Impe, K. (2015, March 26). *How STIX, TAXII and CyBox can help with standardizing threat information*. Security Intelligence. https://securityintelligence.com/how-stix-taxii-and-cybox-can-help-with-standardizing-threat-information/

Vavra, S. (2019, October 22). *Why did cyber command back off its recent plans to call out North Korean hacking?* Cyber Scoop. www.cyberscoop.com/cyber-command-north-korea-lazarus-group-fastcash/

Wagner, T., Mahbub, K., Palomar, E., & Abdallah, A. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security*, 87.

Warren, E. (2018). *Bad credit: Uncovering Equifax' failure to protect Americans' personal information*. Office of Senator Elizabeth Warren. www.warren.senate.gov/files/documents/2018_2_7_%20Equifax_Report.pdf

Weiss, N. E. (2015, June 3). *Legislation to facilitate cybersecurity information sharing: Economic analysis*. Congressional Research Service. No. R43821.

Wendt, D. (2019a). Addressing both sides of the cybersecurity equation. *Journal of Cyber Security and Information Systems*, 7(2).

Wendt, D. (2019b). *Exploring the strategies cybersecurity specialists need to improve adaptive cyber defenses within the financial sector: An exploratory study* [unpublished doctoral dissertation]. Colorado Technical University.

Wisniewski, C. (2020, January 23). *Looking for silver linings in the CVE 2020-0601 crypto vulnerability*. Naked Security. https://nakedsecurity.sophos.com/2020/01/23/looking-for-silver-linings-in-the-cve-2020-0601-crypto-vulnerability/

World Economic Forum. (2019). Global risks report. www.weforum.org/reports/the-global-risks-report-2019

Zheng, D., & Lewis, J. (2015). *Cyber threat information sharing*. Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/publication/150310_cyberthreatinfosharing.pdf

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

4

De-escalation Pathways and Disruptive Technology

Cyber Operations as Off-Ramps to War

Brandon Valeriano and Benjamin Jensen

1 INTRODUCTION

The cyber war long promised by pundits has yet to arrive, failing to match the dramatic predictions of destruction many have been awaiting. Despite fears that digital death is on the horizon (Clarke & Knake, 2014), the international community has seen little evidence. While cyber operations have been used in concert with conventional military strikes from Ukraine (Kostyuk & Zhukov, 2019) to operations against the Islamic State (Martelle, 2018), they have focused more on intelligence collection than shaping direct interdiction. Worst-case scenario nuclear-grade cyberattacks (Straub, 2019) are unlikely and counterintuitive to the logic of cyber action in the international system (Borghard & Lonergan, 2017) where most operations to date tend to reflect political warfare optimized for digital technology, and deniable operations below the threshold of armed conflict (Jensen, 2017; Valeriano et al., 2018).

Decades of research in the field of cybersecurity have laid bare two findings so far: (1) We have failed to witness the death and destruction (Rid, 2020; Valeriano & Maness, 2015) that early prognosticators predicted and (2) digital conflict is typically not a path toward escalation in the international system (Valeriano et al., 2018). Based on survey experiments, when respondents were put in a situation where they had to respond to a militarized crisis using a wide range of flexible response options, more often than not cyber response options were chosen to de-escalate conflicts (Jensen & Valeriano, 2019a, 2019b).

Beyond their raw potential, emergent capabilities like cyber operations are just one among many factors that shape the course of strategic bargaining (Schneider, 2019). New technologies often lead more to questions of resolve and human psychology than objective power calculations about uncertain weapons. The uncertainty introduced by new strategic options, often called exquisite capabilities and offsets, can push states toward restraint rather than war. While these capabilities can certainly lead to dangerous arms races and future risks (Craig & Valeriano, 2016), they tend to play less of an escalatory role in more immediate crisis bargaining. This finding follows work on nuclear coercion in which even nuclear weapons often fail

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

to alter calculations during crises, or have little effect on the overall probability of a crisis (Beardsley & Asal, 2009a, 2009b; Sechser & Fuhrmann, 2017).

How do cyber security scholars explain the evident restraint observed in the cyber domain since its inception (Valeriano & Maness, 2015)? Why have the most powerful states, even when confronted with conventional war, avoided cyber operations with physical consequences? Is it fear or uncertainty that drives the strategic calculus away from escalation during cyber conflicts?

In this chapter, we unpack the strategic logic of interactions during a crisis involving cyber capable actors. We outline the limits of coercion with cyber options for nation-states. After proposing a theory of cyber crisis bargaining, we explore evidence for associated propositions from survey experiments linked to crisis simulations, and a case study of the US-Iranian militarized dispute in the summer of 2019.

2 TOWARD CYBER PEACE AND STABILITY

We are now a field in search of a theory, a theory of cyber peace that explains why cyber capabilities and digital technology offer stabilizing paths in the midst of crisis interactions (Valeriano & Maness, 2015). When we refer to cyber peace, we do not mean the absence of all conflict or positive peace (Roff, 2016), what we have in mind is rather a more measured statement that, while cyber conflicts continue to proliferate, their severity and impact will remain relatively minor (Valeriano & Maness, 2015; Valeriano et al., 2018). This vision of negative peace assumes that violence will continue in the system, but we offer the perspective that during strategic bargaining, cyber options may provide a path toward de-escalation. Cyber operations have the potential to stabilize crisis interactions between rival states. This finding is especially important given that most state-based cyber antagonists are also nuclear armed states (Pytlak & Mitchell, 2016).

On the road to war a state faces many choices regarding the utilization of force and coercion (Schelling, 1960, 1966). Seeking to compel an adversary to back down, a state attempts to display credibility, capability, and resolve (Huth, 1999). To avoid outright conflict, a state can dampen the crisis by making moves that avoid conflict spirals. Much akin to the logic of tit-for-tat struggles of reciprocity (Axelrod & Hamilton, 1981), evidence suggests that actors may choose digital operations to proportionally respond to aggression.

Here we explore the role of cyber operations in producing crisis off-ramps that can stabilize interactions between rival states. That is, during a crisis a state actor is faced with response options to either escalate the conflict, deter further violence, de-escalate the situation, or do nothing. This choice is especially acute during interactions with rivals where tensions are higher. A cyber off-ramp is a strategic choice to either respond in kind, or to de-escalate during a crisis by launching a cyber operation that helps a state set favorable bargaining conditions without losing

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

a significant strategic advantage. By demonstrating weak signals and commitment to the issue at stake, crisis actors can seek to leverage information effects to forestall further escalation.

Cyber operations are not clear paths to peace, but in the context of more dramatic options digital technologies can lead us down a road away from war. During crisis situations, digital technologies can push states away from the brink of escalation by mitigating risks and revealing information to adversaries that helps to manage escalation risks.

3 WHEN DO CRISES ESCALATE?

There is well-established literature on international crises and escalation dynamics, that grew out of the Cold War, which analyzes great power competition as a bargaining process (Schelling, 1958, 2020; Fearon, 1995; Powell, 2002). Conflict as a process is the result of a strategic interactions in which participants attempt to gain an advantage short of the costly gamble of war (Fearon, 1995). During a crisis, each side attempts to signal its capabilities and resolve to the other through deploying military forces, conducting a show of force, making credible threats, and leveraging nonmilitary instruments of power like sanctions and diplomatic demarches.

In this delicate dance, most leaders look to preserve their flexibility to manage escalation risks against the probability of achieving their political objectives. Work on international crises and militarized disputes illustrates this posture through a demonstrated preference for reciprocation strategies in which states adopt a proportional response to threats as a means of maximizing their position short of escalation (Axelrod & Hamilton, 1981; Braithwaite & Lemke, 2011).

Yet, the uncertainty and pressure of a crisis, along with preexisting factors shaping strategic preferences, can pull statesmen away from prudence to the brink of war. States that are rivals are prone to arms races and place a high premium on gaining an advantage in a crisis increasing the probability of escalation (Vasquez, 1993; Sample, 1997; Valeriano, 2013). Territorial disputes tend to be particularly intractable and prone to escalation, especially when there is a recurring history of disputes (Vasquez & Henehan, 2010; Toft, 2014; Hensel & Mitchell, 2017).

Misperception looms large, causing signals to be misinterpreted (Jervis, 2017). Shifts in military capabilities can trigger different risk appetites as the offense–defense balance shifts (Jervis, 1978). There is an open debate about the extent to which espionage and subterfuge in cyberspace alters the security dilemma (Buchanan, 2016). Some work argues that cyber is the perfect weapon and will redefine warfare (Kello, 2017), while other assessments contend it creates a new stability–instability paradox (Lindsay & Gartzke, 2018). Rather than increasing the risk of escalation, cyber operations could act as a crisis management mechanism allowing decision makers to make sharp distinctions between the physical and digital worlds and build active defenses on networks (Libicki, 2012; Jensen & Valeriano, 2019a; Valeriano & Jensen, 2019).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

4 THE LOGIC OF CYBER OFF-RAMPS

This chapter helps develop a midrange theory hypothesizing that cyber operations are a possible mechanism for helping states manage crises in a connected world.

First, in crisis settings between rival states cyber operations are best thought of as a coercive capability (Borghard & Lonergan, 2017). In addition to their value in intelligence operations (Rovner, 2019), they allow states to disrupt and degrade rival networks.

As instruments of coercion, cyber operations tend to produce fleeting and limited effects, best characterized as ambiguous signals (Valeriano et al., 2018). Ambiguous signals are "covert attempts to demonstrate resolve that rely on sinking costs and raising risks to shape rival behavior" (Valeriano et al., 2018, p. 13). States engage in covert communication, probing each other during a crisis (Carson, 2020). The benefit of cyber operations is that they are a weak signal that can be denied, preserving bargaining space while still demonstrating a willingness to act. This makes cyber operations a low cost, low payoff means of responding early in a crisis.

Second, experimental studies show that the public tends to treat cyber operations different than they do other domains. There are also key threshold dynamics associated with cyber operations. In a recent study, Kreps and Schneider (2019) found that "Americans are less likely to support retaliation with force when the scenario involves a cyberattack even when they perceive the magnitude of attacks across domains to be comparable." For this reason, cyber operations offer a means of responding to a crisis less likely to incur domestic audience costs that could push leaders to escalate beyond their risk threshold.

Avoiding escalation is especially appealing since there are indications that most twenty-first century great powers maintain a public aversion to casualties. Even authoritarian regimes limit reporting and use a mix of private–military companies and proxies to hide the true cost of war from their citizens (Reynolds, 2019). Given this emerging dynamic, cyber operations offer states a means of responding to a crisis without triggering direct, immediate human costs that can often lead to an emotional, as opposed to a rational, conflict spiral. Cyber operations help states manage thresholds in crisis interactions.

Third, and less explored by the cyber security literature to date, cyber operations are defined by unique substitutability dynamics. To say cyber operations are subject to substitution effects implies that states evaluate the trade-offs inherent in using cyber instruments when signaling another state.

In economics, there is a long history of using marginal analysis (Marshall, 1890; Krugman et al., 2008) to evaluate trade-offs in production and consumption. In microeconomics, the marginal rate of substitution is the extent to which a consumer will give up one good or service in exchange for another (Krugman & Wells, 2008). The two goods or services, even courses of action, can be perfect substitutes, in which case they are interchangeable, or imperfect substitutes – in which case the

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

indifference curve shifts. Furthermore, there is a distinction between within-group and crosscategory substitution in economics and psychological studies of consumer choice (Huh et al., 2016). There is also a long history of work on foreign policy substitutability in international relations (Most & Starr, 1983; Starr, 2000; Most & Starr, 2015). This research maps out when similar acts, as substitutes, trigger different (Palmer & Bhandari, 2000) or similar foreign policy outcomes (Milner & Tingley, 2011).

Applied to contemporary escalation and foreign policy, contemporary leaders evaluate whether to substitute a cyber effect for a more conventional instrument of power. We propose that there are unique substitutability dynamics involved with selecting cyber operations during strategic bargaining episodes. If cyber operations are not efficient substitutes, then they require an increased number or complements. To the extent that cyber operations are an imperfect substitute, a state would have to use more cyber effects to compel an adversary than, for example, traditional diplomatic demarches or threats of military action. The central question for decision makers thus concerns the ideal typical crosselasticity of demand for cyber operations.

We theorize that cyber operations are subject to certain characteristics that make them weak substitutes, and better thought of as complements. In microeconomics, a complement implies the use of one good or service that requires the use of another complementary good or service. If you use a printer, you are going to need a constant supply of toner and paper. With respect to cyber operations, it means that, as shaping mechanisms, they will tend to be paired with at least one more instrument of power to compensate for their weak substitutability as an ambiguous signal subject to threshold effects. This logic follows earlier findings that states will tend to use cyber operations in conjunction with other instruments of power that include both positive and negative inducements (Valeriano et al., 2018).

Two additional dynamics alter the elasticity of demand for cyber effects in crisis bargaining. First, the elasticity of demand is skewed by the dual-use dynamic of cyber operations. Cyber operations tend to be a use and lose capability limiting when states will risk employing high-end capabilities (Jensen & Work, 2018). Leaders who have cyber probes spying on adversary systems worry about sacrificing their digital scouts for fleeting attack opportunities, a calculation known in US Joint doctrine as intelligence gain/loss.[1] They also worry about burning capabilities by exposing their operations. Many cyber capabilities can be both intelligence and tools of subterfuge simultaneously. A tool kit used to access a rival states computer networks and extract information can also be used to deliver malicious code.

[1] See JP 3-12 Cyber Operations: www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_12.pdf. Of note, at the apex of national security, decision makers also weigh political gain/loss (PGL) and technical gain/loss (TGL).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Back to the concept of substitution, this dynamic means that states must pay information costs to burn access and deliver their payload. Once you attempt to achieve an effect beyond espionage, one increases the risk that the rival state knows you are accessing their networks. Information costs and the opportunity cost of future intelligence lost to achieve a cyber effect skew elasticity and lowers escalation risks. When a state does employ cyber capabilities to respond to a crisis scenario, they will prefer lower end capabilities to reduce information costs. There are unlikely to employ more exquisite tools to achieve a cyber fait accompli that produces an escalation spiral. More importantly, they will look for specific conditions to use cyber substitutes, such as when a rival state has less cyber capability and thus reduces information costs associated with burning a digital spy.

Second, the elasticity of demand is further skewed by a second category of information cost, the shadow of the future (Axelrod, 1984; Axelrod & Keohane, 1985). States like the United States have more than one rival, and even when a state has a single rival they expect to interact with them in the future. Therefore, burning a tool or tool kit in the present risks losing that capability relative to either another rival in the present or a target state in the future. This compounds the information costs that skew the indifference curve. As a result, cyber operations will tend to be used as complements, combined with other instruments of power to increase the expected marginal effect. They can be used as substitutes, but only under conditions where states assess a lower likelihood of paying additional information costs associated with the dual-use dimension and shadow of the future. On its own, the extent to which a cyber operation is substitutable could trigger a security dilemma (Herz, 1950; Glaser, 1997; Booth & Wheeler, 2007).[2] Yet, the substitution of cyber capabilities occurs in a larger context defined by ambiguous signals and threshold effects that dampen escalation risks. These properties help states escape the security dilemma and view cyberattacks as less escalatory than conventional military operations. In the end, cyber capabilities are weak substitutes and will be used more as complements to manage escalation outside of narrow conditions.

Taken together, the above logic of weak coercive potential, thresholds, and substitution effects produces the following three hypotheses.

H1. Cyber operations are not escalation prone.

Observations from cases and survey experiments should demonstrate that when cyber capabilities are present they are not associated with increased escalation. The null hypothesis is that cyber operations are associated with escalation spirals. The hypothesis is better evaluated through large-N methods associated with either past, observed cyber incidents or survey experiments examining escalation preferences when compared actively to the use of other instruments of power. Case studies

[2] Blue networks are home networks, gray networks are unallied network spaces, and red networks are opposition systems.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

would show more the process and sequence associated with using cyber operations. One would expect to see cyber instruments used to check escalation as a weak, proportional alternative before crossing into higher thresholds.

H2. Cyber operations are more likely to be used as complements when states consider escalating a crisis.

Due of their weak substitutability, cyber operations will tend to complement other instruments of power. There are inherent cross-domain effects associated with modern crisis management (Gartzke & Lindsay, 2019). When examining survey experiments on crisis decision making involving selecting between cyber and noncyber response options, there should more instances of combining cyber effects with other instruments of power. The null hypothesis would be that there is no relationship between cyber escalation and using multiple instruments of power.

H3. Cyber operations are more likely to be used as substitutes for other measures of power when there are no indications of rival cyber activity.

Since cyber operations tend to be weak substitutes, due to information costs and the elasticity of demand, there should be narrow scope conditions that shape when and how they are used in place for more traditional instruments of power. The state will want to minimize the shadow of the future and avoid losing the inherent value of cyber capabilities that are unknown to the adversary. This dynamic implies that in survey experiments one would expect to see a higher percentage use of cyber tools in treatments where there are no indications the adversary is using cyber operations. This initial indication helps respondents gauge the substitutability costs and inherent trade-offs of using cyber capabilities.

5 HOPE AMONGST FEAR: INITIAL EVIDENCE

5.1 *Research Design*

Demonstrating that cyber operations can serve as crisis off-ramps and represent a common strategic choice to respond proportionally during crisis interactions can be a difficult proposition. The goal is to find evidence, under a controlled setting, when a state will have to make a choice between an option that might cause significant damage, an option that will cause little or no harm, the option of doing nothing, and the ability to wage a cyber operation against the opposition.

We propose two methods to investigate our propositions, a theory-guided case study investigation and a survey experiment using crisis simulations and wargames. Once the plausibility of our propositions is determined, we can follow-up our examinations with further support and evidence through follow on experiments. This is not a simple process and we only begin our undertaking here.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

The case study presented here represents a theory-guided investigation according to Levy's (2008) typology. These case studies are "structured by a well-developed conceptual framework that focuses attention on some theoretically specified aspects of reality and neglects others" (Levy, 2008, p. 4). In these cases, we cannot rule out other theoretical propositions for the cause of de-escalation, but can demonstrate the process of how cyber activities provide for off-ramps on the road to conflict.

Such case studies can also serve as plausibility probes. According to Eckstein (1975, p. 108), plausibility probes "involve attempts to determine whether potential validity may reasonably be considered great enough to warrant the pains and costs of testing." We can only pinpoint the impact of a cyber operation as a choice and examine the outcome – de-escalation during a case study investigation.

Case studies are useful, but do not provide controlled situations where there are clear options and trade-offs for leadership. It might be that a cyber option was decided before the crisis was triggered, or that a cyber option in retaliation was never presented to the leader. Here, we will use a short case study to tell the story of how a cyber operation was chosen and why it represented a limited strike meant to de-escalate a conflict, but will pair this analysis with an escalation simulation.

Deeper investigations through proper controlled settings can be done through experimental studies. In this case, experimental wargames where a group of actors playing a role must make choices when presented with various options. Our other option is survey experiments to demonstrate the wider generalizability of our findings, but such undertakings are costly and time intensive.

Experiments are increasingly used in political science to evaluate decision making in terms of attitudes and preferences (Hyde, 2015; Sniderman, 2018). While there are challenges associated with external validity and ensuring that the participants reflect the elites under investigation, experiments offer a rigorous means of evaluating foreign policy decision making (Renshon, 2015; Dunning, 2016). For the experiment below, we employ a basic 2 × 2 factoral design.

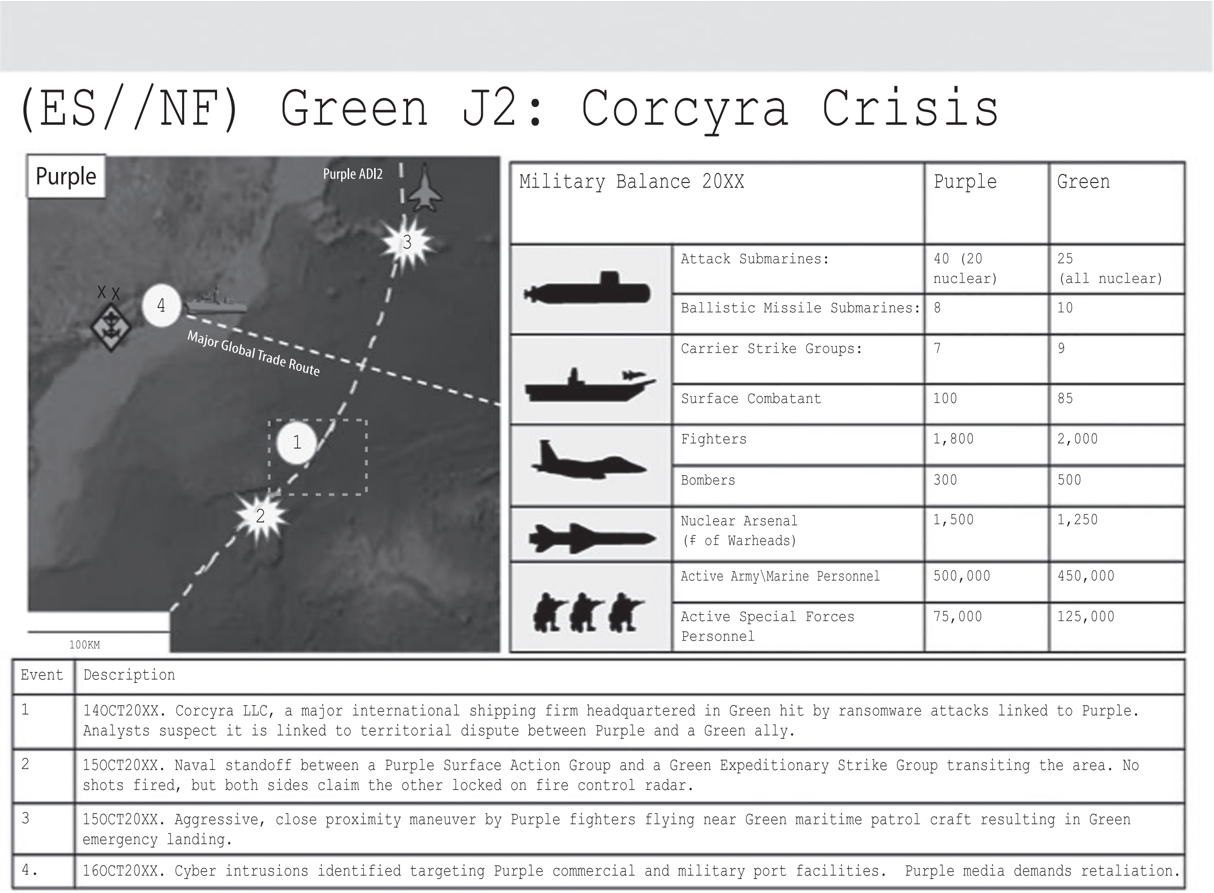
5.2 *Wargames as Experiments*

To date, research on cyber operations have focused either on crucial case studies (Lindsay, 2013; Slayton, 2017), historical overviews (Healey & Grindal, 2013; Kaplan, 2016), and quantitative analysis (Valeriano & Maness, 2014; Kostyuk & Zhukov, 2019; Kreps & Schneider, 2019). Recently, researchers have expanded these techniques to include wargames and simulations analyzed as experiments.

There is a burgeoning literature on the utility of wargames and simulations for academic research. Core perspectives generally define the purpose and utility of wargames, failing to include the wider social science implications of new methodologies defaulting toward the perspective that war-gaming is an art (Perla, 1990; Van Creveld, 2013). More recently, there has been an increasing amount of research offering

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC39806486 7A9B5EA938D



(ES//NF) Green J2: Corcyra Crisis

Military Balance 20XX		Purple	Green
	Attack Submarines:	40 (20 nuclear)	25 (all nuclear)
	Ballistic Missile Submarines:	8	10
	Carrier Strike Groups:	7	9
	Surface Combatant	100	85
	Fighters	1,800	2,000
	Bombers	300	500
	Nuclear Arsenal (# of Warheads)	1,500	1,250
	Active Army\Marine Personnel	500,000	450,000
	Active Special Forces Personnel	75,000	125,000

Event	Description
1	14OCT20XX. Corcyra LLC, a major international shipping firm headquartered in Green hit by ransomware attacks linked to Purple. Analysts suspect it is linked to territorial dispute between Purple and a Green ally.
2	15OCT20XX. Naval standoff between a Purple Surface Action Group and a Green Expeditionary Strike Group transiting the area. No shots fired, but both sides claim the other locked on fire control radar.
3	15OCT20XX. Aggressive, close proximity maneuver by Purple fighters flying near Green maritime patrol craft resulting in Green emergency landing.
4.	16OCT20XX. Cyber intrusions identified targeting Purple commercial and military port facilities. Purple media demands retaliation.

FIGURE 4.1 *Diagram from Wargame Simulation.*

a social science perspective on war-gaming as a research methodology (Schneider, 2017; Pauly, 2018; Jensen and Valeriano, 2019a, 2019b). The perspective that wargames can add to our knowledge about crisis bargaining under novel technological settings is one we follow herein (Reddie et al., 2018; Lin-Greenberg et al., 2020).

To evaluate the utility of cyber operations in a crisis, the researchers used a conjoint experiment linked to a tabletop exercise recreating national security decision making. Small teams were given packets that resembled briefing materials from US National Security Council (NSC) level deliberations based on guidance from NSC staffers from multiple prior administrations. The packets outlined an emerging crisis between two nucleararmed states: Green and Purple. The graphics and descriptions tried to obscure the crisis from current states, such as China and the United States. The respondents were asked to nominate a response to the crisis, selecting from a range of choices capturing different response options using diplomatic, information, military, and economic instruments of power. Each instrument of power had a scalable threshold of options, from de-escalatory to escalatory. This range acted as a forced Likert scale. Figure 4.1 shows a sample page from the respondent packets outlining the road to crisis and balance of military capabilities.

The packets were distributed to a diverse, international sample of 400 respondents in live session interactions. In the terms of the types of respondents who participated, 213 were students in advanced IR/political science classes, indicative of individuals likely to pursue a career in foreign policy, 100 were members of the military with the most common rank being major (midcareer), 40 were members of a government involved with foreign policy decision-making positions, 19 were involved with major international businesses, and 13 opted not to disclose their occupation, while 15 left it blank. Of these respondents there were 267 male respondents, 110 female respondents, and 4 who preferred not to say, while 19 opted to leave it blank.[3] With respect to citizenship, 295 respondents were US citizens, 87 were non-US citizens, and 4 preferred not to say, while 14 left their response blank.[4]

These participants were randomly assigned to one of four treatment groups:

Scenario 1. A state with cyber response options (cyber resp) that thinks the crisis involves rival state cyber effects (cyber trig);

Scenario 2. A state with no cyber response options (no cyber resp) that thinks the crisis involves rival state cyber effects (cyber trig);

Scenario 3. A state with cyber response options (cyber resp) that thinks the crisis does not involve rival state cyber effects (no cyber trig); and

Scenario 4. A state with no cyber response options (no cyber resp) that thinks the crisis does not involve rival state cyber effects.

[3] Participants were encouraged to identify gender based on preference and leave it blank if they were gender fluid in most settings to create a safe, inclusive environment.

[4] Participants were encouraged to fill out this option only if they felt comfortable to preserve maximum anonymity and create a safe, inclusive space.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 4.1 *Treatment groups*

Treatment			Number
1.	Cyber Response Options (Yes)	Assumed Rival Cyber Activity (Yes)	100
2.	Cyber Response Options (No)	Assumed Rival Cyber Activity (Yes)	100
3.	Cyber Response Options (Yes)	Assumed Rival Cyber Activity (No)	100
4.	Cyber Response Options (No)	Assumed Rival Cyber Activity (No)	100

N = 400.

These treatments allowed the researchers to isolate cyber response options and assumptions about the role of rival state cyber effects in the crisis. These treatment groups are listed in Table 4.1.

To measure escalation effects associated with cyber capabilities (H1), the survey experiment examined participant response preferences using the respondent initial preference (RESP) variable. This variable asked the survey respondents to indicate their initial reaction and preferred response to the crisis as de-escalate (1), adopt a proportional response (2), escalate (3), or unknown at this time (4). Coding along these lines allowed the researchers to factor in uncertainty and capture if there were any differences between what the survey respondents wanted to do initially, and what they selected to do after reviewing approved response options across multiple instruments of power. Furthermore, as a 2 × 2 experiment focused on attitudes and preferences, the RESP variable helped the team determine if the four different treatments altered the decision to escalate as a cognitive process, and how each participate viewed their options given limited information in a rivalry context. The results are shown in the contingency table (Table 4.2 and Figure 4.2).

Escalation was generally low with only twenty respondents preferring escalation. When they did opt to escalate, neither the presence of cyber response options nor the adversary use of cyber seemed to affect their response preference. Alternatively, when states had cyber response options and there were no signs of rival state cyber effects, participants opted to de-escalate (57) more than expected (47.5). The results were inverse when states were in a crisis that lacked cyber options and adversary cyber effects (treatment 4). Here there were less observed preferences to de-escalate (28) than expected (42.5) and more instances of proportional responses (67) than expected (49.8). The results also lend themselves to categorical variable tests for association using the phi coefficient (Sheskin, 2020). The phi coefficient is 0 when there is no association and 1 when there is perfect association. The value is .286 indicating a weak but significant relationship between the treatment group and escalation preferences consistent with the hypothesis. Cyber options were not associated with escalation and were, in fact, linked to preferences for de-escalation.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 4.2 *Contingency results by treatment*

			Treatments				
			Cyber Trig Cyber Resp	Cyber Trig No Cyber Resp	No Cyber Trig Cyber Resp	No Cyber Trig No Cyber Resp	Total
RESP	De-escalate	Count	41	44	57	28	170
		Expected Count	42.5	42.5	42.5	42.5	170.0
		% within RESP	24.1	25.9	33.5	16.5	100.0
		% within SCENARIO	41.0	44.0	57.0	28.0	42.5
		% of Total	10.3	11.0	14.2	7.0	42.5
		Standardized Residual	−.2	.2	**2.2	**−2.2	
	Proportional	Count	51	46	35	67	199
		Expected Count	49.8	49.8	49.8	49.8	199.0
		% within RESP	25.6	23.1	17.6	33.7	100.0
		% within SCENARIO	51.0	46.0	35.0	67.0	49.8
		% of Total	12.8	11.5	8.8	16.8	49.8
		Standardized Residual	.2	−.5	**−2.1	**2.4	
	Escalate	Count	5	3	7	5	20
		Expected Count	5.0	5.0	5.0	5.0	20.0
		% within RESP	25.0	15.0	35.0	25.0	100.0
		% within SCENARIO	5.0	3.0	7.0	5.0	5.0
		% of Total	1.3	0.8	1.8	1.3	5.0
		Standardized Residual	.0	−.9	.9	.0	

(continued)

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 4.2 *(continued)*

		Treatments				
		Cyber Trig Cyber Resp	Cyber Trig No Cyber Resp	No Cyber Trig Cyber Resp	No Cyber Trig No Cyber Resp	Total
Uncertain	Count	3	7	1	0	11
	Expected Count	2.8	2.8	2.8	2.8	11.0
	% within RESP	27.3	63.6	9.1	0.0	100.0
	% within SCENARIO	3.0	7.0	1.0	0.0	2.8
	% of Total	0.8	1.8	0.3	0.0	2.8
	Standardized Residual	.2	**2.6	–1.1	–1.7	
Total	Count	100	100	100	100	400
	Expected Count	100.0	100.0	100.0	100.0	400.0
	% within RESP	25.0	25.0	25.0	25.0	100.0
	% within SCENARIO	100.0	100.0	100.0	100.0	100.0
	% of Total	25.0	25.0	25.0	25.0	100.0

$X^2 = 32.723$, $p < .000$ (two-sided), ** = standardized residual is ±1.96.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC39806486 7A9B5EA938D

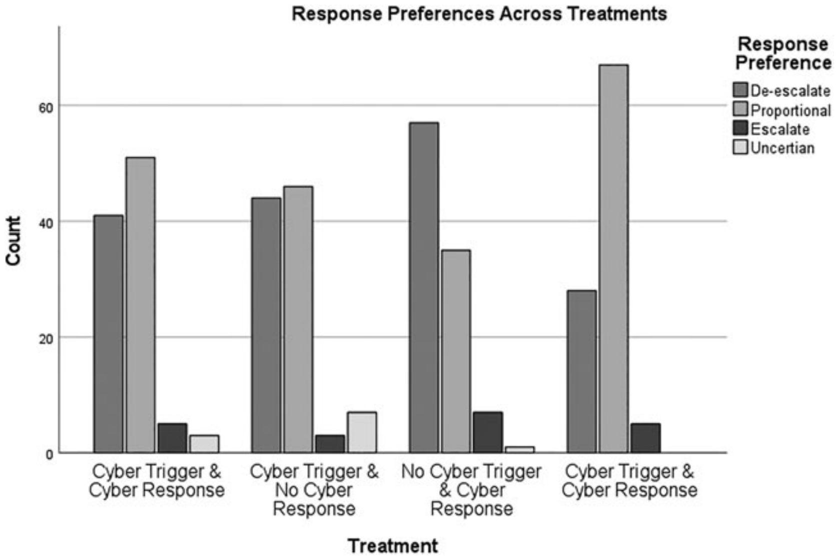


FIGURE 4.2 *Response preferences from wargame simulation.*

A second measure of escalation allows the team to differentiate between the RESP and the overall degree of potential escalation based on the instruments of power selected. This measure is less effective since it does not capture the attitude and preference as a cognitive process in line with best practices in experiments, but does allow the researchers to further triangulate their findings. The researchers created a variable odds of escalation (OES) and average odds of escalation (OESAAVG). OES is a summation and adds the escalation scores from across the actual response options selected. OESAAVG is a binary variable coded 1 if the OES score is over the average and 0 if it is under the average (Table 4.3). OESAAVG allows the researchers to look across the treatments and see if there are differences when cyber response options are present and absent.

The results cast further doubt on cyber operations as being escalatory. Both treatments 1 and 3 had less combined instruments of power above the average coercive potential (29, 30) than expected (37, 37). Of particular interest, when states had cyber response options and escalated, the magnitude tended to be less with treatment 1 seeing 29 instances of above average coercive potential versus 37 expected (–1.3 standardized residual) and treatment 3 seeing 30 instances versus 37 expected (–1.2 standardized residuals). These contrast with treatment 2 where there is a cyber trigger and no cyber response options available. Here there were 48 instances of above average coercive potential versus 37 expected (1.8 standardized residual). Cyber appears to have a moderating influence on how participants responded to the crisis.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 4.3 *Expected count of escalation events*

			SCENARIO				
			1	2	3	4	Total
OESAAVG	0	Count	71	52	70	59	252
		Expected Count	63.0	63.0	63.0	63.0	252.0
		Standardized Residual	1.0	−1.4	.9	−.5	
	1	Count	29	48	30	41	148
		Expected Count	37.0	37.0	37.0	37.0	148.0
		Standardized Residual	−1.3	1.8	−1.2	.7	
Total		Count	100	100	100	100	400
		Expected Count	100.0	100.0	100.0	100.0	400.0

$X^2 = 10.725$, $p < .013$ (two-sided), ** = standardized residual is ±1.96.

Turning to the second hypothesis, to measure complementary effects associated with the survey experiment, the researchers examined how participants combined instruments of power. Participants were allowed to recommend three response options to the crisis. These response options were organized by instruments of power on the aforementioned Likert scale. Each instrument had six options. In treatments where participants had cyber response options, six additional options were added each with an equivalent level of escalation. This gave participants a total of twelve responses in cyber treatments. Since the packets involved four instruments of power (diplomatic, information, military, economic), participants had a total of 24 response options in noncyber treatments (treatments 2, 4) and 48 in cyber response treatments (1, 3). Participants could choose three response options all in one instrument of power, or spread them across multiple instruments of power. Table 4.4 shows the number of response options selected for each instrument of power across the treatments below. There were no statistically significant differences across the treatments with respect to the distribution of the responses.

In each survey experiment, the researchers used this information to create a variable called COMB (combined) that measured the number of instruments of power a respondent used. This number ranged from one to three. Since the survey experiments asked participants to select three options, they could either select three options from any one instrument of power or employ up to three combined instruments of power. To confirm the second hypothesis, one would need to see a higher than expected instances of combining instruments of power comparing conventional versus cyber escalation preferences.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 4.4 *Treatment groups and instrument of power response preferences*

Treatment	Diplomatic	Information	Military	Economic
1	80	88	57	53
2	81	84	54	67
3	70	85	77	50
4	71	86	60	62

$X^2 = 12$, $p < .213$ (two-sided).

TABLE 4.5 *Conventional versus cyber escalation*

	Conventional Escalation		Cyber Escalation	
Inst Power	No Escalation	Escalation	No Escalation	Cyber Escalation
1	+0(.5)	+1(.5)	6(6.4)	+1(.6)
2	18(16.8)	15(16.2)	19(23.8)	**7(2.2)
3	84(84.7)	82(81.7)	158(152.8)	9(14.2)
	$X^2 = 1.217$, $p < .544$ (two-sided) N = 200 (Treatments 2, 4)		$X^2 = 13.726$, $p < .005$ (two-sided) N = 200 (Treatments 1, 3)	

** = standardized residual > 1.96.

+ = count is less than 5 (cannot evaluate).

To evaluate hypothesis two along these lines, the researcher separated treatments 2 and 4 and 1 and 3 to compare escalation preferences and combined instruments of power. In Table 4.5, the conventional escalation column shows how many times respondents used 1, 2, or 3 instruments of power, differentiating between treatments that saw escalation and no escalation.[5]

Third, to evaluate substitution, the researchers compare percentages. There should be a higher rate of substitution, measured as using a cyber option, in treatment 3 than in treatment 1. In treatment 3, participants have no evidence the rival state is using cyber capabilities thus making them more likely to substitute cyber effects due to the lower, implied information costs. A respondent would look at the situation and see more utility in using cyber because no adversary cyber effects are present. Alternatively, when adversary cyber effects are present, participants will assess higher information costs. They will be more

[5] For this test, the escalation measure was the coercive potential and whether any instrument selected was greater than 3 on the previously discussed Likert scale for each instrument of power.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 4.6 *Coercive potential*

Treatment	Escalation	Escalation Involved Cyber
1	35	6 (17.14%)
2	50	NA
3	21	11 (52.38%)
4	48	NA

N = 400.

TABLE 4.7 *Coercive potential and cyber substitution*

Treatment	Diplomatic	Information	Military	Economic
1	20(3)	10(4)	12(1)	7(1)
3	10(5)	7(5)	14(7)	4(2)

N = 2,000.

concerned about adversaries being able to mitigate the expected benefit of any cyber response (Table 4.6).

As predicted, there was more observed substitution in treatment 3, as opposed to treatment 1. In treatment 3, 52.38% of the response options selected (i.e., coercive potential) involved cyber equivalents compared with 17.14% for treatment 1. Because there were no indications of adversary cyber capabilities in this treatment, participants likely perceived a cross-domain advantage, hence less information costs. This alters the hypothetical elasticity of demand making cyber a more perfect substitute. Table 4.7 breaks out the substitution further.

In treatment 1, cyber responses were substituted at a higher rate for information effects (40%) than other instruments of power. Three of the four substitutions involved the option to "burn older exploits in adversary systems disrupting their network operations in order to signal escalation risks."

In treatment 3, cyber responses were heavily used to substitute for conventional responses over 50% of the time. The most common military substitution (4/7) involved opting to "compromise data of individual members of the military to include identify theft, fraud, or direct social media messaging." This option substituted for the conventional response: "Conduct a public show of force with air and naval assets challenging known defense zones and testing adversary response." Participates opted for information warfare, or more conventional displays of military force. The most common information substitution remained burning "older exploits in adversary systems disrupting their network operations in order to signal escalation risks." The most common diplomatic substitution in the packet was "use spear phishing, waterholing,

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

and other methods to expose sensitive political information." Again, information warfare was a substitute for more conventional forms of coercion when the adversary posture suggests a low probability of response to information operations.

Another factor stands out when looking at the descriptive statistics associated with differentiating conventional and cyber escalation, measured as coercive potential. As seen in Table 4.6, there is a higher observed rate of coercive potential in noncyber response treatments. The available of cyber response options appears to reduce the coercive potential by substituting information warfare for more traditional approaches to coercion.

Overall, we have evidenced that cyber response options can moderate a conflict between rival powers. Respondents generally used cyber options to either respond proportionally or seek to de-escalate the situation until more information can be gathered. What we cannot explain is whether or not the results were influenced by the presence of nuclear weapons on both sides, different regime types, and other possible confounding variables because our sample was not large enough to enable additional treatments.

6 Case Study Probe: The United States and Iran

To further examine the concept of cyber off ramps and contemporary escalation dynamics, we turn to a theory-guided case study examination (Levy, 2008). Since survey experiments are prone to external validity challenges (Renshon, 2015), a case analysis helps triangulate the findings from the three hypotheses. To this end, interactions between the United States and Iran in the summer of 2019 offer a viable case for examination (Valeriano & Jensen, 2019). Referring to the prior hypotheses, we argue that cyber operations are not escalation prone (H1). We also note that cyber operations are more likely to be used as complements when states do consider escalating (H2), and that cyber operations are more likely to be used as substitutes when there are no indications of rival cyber activity (H3). We now examine our developing theory's plausibility in the context of this case.

6.1 *Origins*

The full picture of what happened between Iran and the United States in the summer of 2019 will continue to develop as classified information is released, but what we do know suggests there was a significant confrontation with cyber operations playing a role as a coercive instrument alongside diplomatic, economic, and military inducements in the dispute. Given that Iran and the United States maintain an enduring rivalry and have a history of using force, even if through proxies, this case was particularly escalation prone. Yet, instead of going to war, Tehran and Washington pulled back from the brink. The key question is why?

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

As long-term rivals, the United States and Iran have been at loggerheads over the control of the Middle East and resource access for decades (Thompson & Dreyer, 2011). The origins of the contemporary rivalry between Iran and the United States started, from an Iranian perspective, in 1953 when the CIA helped their UK counterparts stage a coup (Kinzer, 2008). From the US perspective, the rivalry dates to the Iranian Revolution and the overthrow of the Shah in 1979, installed in the 1953 coup (Nasri, 1983). The new regime, led by Ayatollah Ruhollah Khomeini, launched a revisionist series of direct and proxy challenges against US interests in the region (Ramazani, 1989) that culminated in a protracted conflict with Iraq. During the Iran–Iraq War, the United States backed Iran's rivals, including Iraq and the larger Gulf Cooperation Council. Iran in turn backed Shiite groups across the Middle East implicated in attacking US forces in the Lebanon.

In the aftermath of the Iranian Revolution and during the subsequent Iran–Iraq War, the United States engaged in limited but direct military engagements with Iran, including the failed Desert One raid to rescue American hostages (1980), and during Operation Earnest Will (1987–1988) in which the US Navy escorted Gulf State oil tankers in a convoy to protect them from Iranian military forces (Wise, 2013). This period included multiple naval skirmishes such as Operational Praying Mantis (1988) and Operational Nimble Archer (1988) in which US forces attacked Iranian oil rigs and military forces in retaliation for Iranian mining in the Strait of Hormuz and repeated attacks. Contemporary US perspectives on Iranian motives and likely foreign policy preferences emerged during this period, with the Washington foreign policy establishment seeing Iran as a revisionist, revolutionary state.[6] Similarly, Iranian attitudes toward the United States hardened even further as Washington labeled the country part of an Axis of Evil (Shay, 2017) and invaded its neighbor, Iraq. Iran opted to counter by funding proxy Shiite groups in Iraq and undermining the transitional Iraqi government.[7]

Parallel to its proxy struggle with the United States in Iraq, Tehran sponsored terror groups that attacked US interests across the region and accelerated its nuclear weapons program.[8] Starting in 2003, the International Atomic Energy Agency started pressuring Iran to declare its enrichment activities, which led to multilateral diplomatic efforts starting in 2004. These efforts culminated in UN Security Council resolutions expanding sanctions on Iran over the subsequent years, and the US joining the multilateral effort (P5+1) in April 2008 following a formal Iranian policy review. Backed by the larger range of diplomatic and economic sanctions that had been in place since the Iranian Revolution, the pressure resulted in the 2015 Joint Comprehensive Plan of Action (JCPOA). This agreement limited Iran's ability to develop nuclear

[6] For an overview of US intelligence estimates during this period, see a 1985 declassified CIA study: www.cia.gov/library/readingroom/docs/CIA-RDP86T00587R000200190004-4.pdf.

[7] This analysis focuses on the context of the dyadic rivalry and does not address the role of Israel and other US security partners in the Middle East, such as Saudi Arabia.

[8] For a timeline of Iranian nuclear efforts and related diplomacy, see the Arms Control Association Timeline (updated September 2020): www.armscontrol.org/factsheets/Timeline-of-Nuclear-Diplomacy-With-Iran.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

weapons and included European allies as treaty members distributing the burden of enforcement internationally (Mousavian & Toossi, 2017).

In 2018, the Trump administration withdrew from the agreement, arguing that Iran was still building nuclear weapons and directing proxy warfare against US allies (Fitzpatrick, 2017). The Trump administration wanted to move past the JCPOA agreement, which had reduced tensions in the region. Instead, the Trump administration ramped up sanctions and designated the Islamic Revolutionary Guard Corps, with the Quds force (Tabatabai, 2020), a terrorist organization in 2019 (Wong & Schmitt, 2019). The leader of the organization, QasemSoleimani, became a prime target (Lerner, 2020).

6.2 *Cyber and Covert Operations*

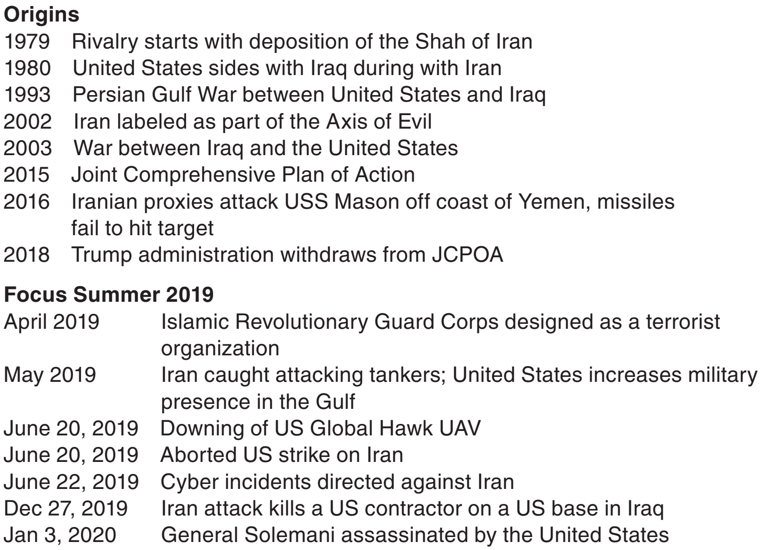
Given Iran's use of proxies, covert operations generally color the relationship between Iran and United States. These activities included the use of cyber capabilities. The United States and Iran were deep in a cyber rivalry, with twenty cyber conflicts between 2000 and 2016 (Valeriano et al., 2018). Data on cyber interactions only begin in 2000, making it difficult to catalog the full range of covert and clandestine activity between 1979 and 2000.

With respect to cyber operations, the United States likely initiated seven cyber operations while Iran launched thirteen (Maness et al., 2019). The most significant event was when the United States and Israel launched the Stuxnet attack, which disabled centrifuges in the Natanz nuclear power plant (Lindsay, 2013). The overall impact of the attack on the Natanz plant is intensely debated, but assessment at the time suggested a limited overall impact on Iran's ability to produce nuclear materials (Barzashka, 2013). It is still unknown what effect the Stuxnet attack had on Iranian internal calculations and assessment of US capabilities.

The pattern between the United States and Iran has often been for the United States to rely on cyber espionage and degrade operations to harm Iranian interests and activities, while Iran generally seeks to avoid direct confrontation in cyberspace (Valeriano & Maness, 2015). Saudi Arabia is a frequent proxy cyber target of Iran, given that the United States is seen as its protector and ally. Iran's actions against the United States mostly entail basic espionage, economic warfare, and the typical probes and feints in cyberspace (Eisenstadt, 2016).

Another key aspect of the covert competition, and the prime threat that Iran offered to the United States, was the use and control of proxy forces in the region. The Iranian Quds force controlled proxy actors in the region (Eisenstadt, 2017), with Houthi forces seeking to attack forces in the region with Scud missiles (Johnston et al., 2020). The awareness that Hezbollah was taking clear direction from Iran altered the dynamics of the dispute between Israel and its regional rivals (Al-Aloosy, 2020). Entering the summer of 2019, Iran's use of proxy forces dominated the concerns of the Trump administration (Simon, 2018; Trump, 2018).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D



Origins

1979	Rivalry starts with deposition of the Shah of Iran
1980	United States sides with Iraq during with Iran
1993	Persian Gulf War between United States and Iraq
2002	Iran labeled as part of the Axis of Evil
2003	War between Iraq and the United States
2015	Joint Comprehensive Plan of Action
2016	Iranian proxies attack USS Mason off coast of Yemen, missiles fail to hit target
2018	Trump administration withdraws from JCPOA

Focus Summer 2019

April 2019	Islamic Revolutionary Guard Corps designed as a terrorist organization
May 2019	Iran caught attacking tankers; United States increases military presence in the Gulf
June 20, 2019	Downing of US Global Hawk UAV
June 20, 2019	Aborted US strike on Iran
June 22, 2019	Cyber incidents directed against Iran
Dec 27, 2019	Iran attack kills a US contractor on a US base in Iraq
Jan 3, 2020	General Solemani assassinated by the United States

FIGURE 4.3 *Iran–United States Case Timeline* (Source) [no date].

6.3 *The Summer 2019 Crisis*

As the summer began in 2019, tensions accelerated due to concerns about Iranian proxy warfare, the use of cyber actions in the region, and the pursuit of nuclear weapons after the end of the JCPOA (see Figure 4.3 for the timeline of events). In addition to increased hacking activities, Iran attacked tankers in the Persian Gulf, with two incidents occurring in May of 2019. At one point, Iranian operatives were seen placing unidentified objects on the hull of a tanker before it was disabled. Iran "called the accusations part of a campaign of American disinformation and 'warmongering'" (Kirkpatrick et al., 2019).

Following intelligence reports that Iran was plotting an attack on US interests in the Middle East on May 5, 2019, National Security Adviser, John Bolton, announced (Bolton, 2019) the deployment of a carrier strike group and bomber task force to the Middle East to "send a clear and unmistakable message to the Iranian regime that any attack on the United States interests or those of our allies will be met with unrelenting force." In response, on May 12 the crisis escalated with four commercial vessels, including two Saudi Aramco ships, targeted by sabotage attacks attributed to Iran in the Gulf of Aden (Yee, 2019). By May 13, the Pentagon announced plans to deploy as many as 120,000 troops in the region in additional fighter squadrons and naval task forces already headed to the region (Schmitt & Barnes, 2019). In response, on May 14 Iranian proxies in Yemen launched a massive attack against Saudi oil infrastructure using a mix of drones and cruise missiles (Hubbard et al., 2019). By the

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

end of May, the United States implicated Iran proxies in firing rockets at US interests in Iraq and responded with additional troop deployments and weapon sales to Saudi Arabia. These measures added to the range of economic sanctions the Trump administration initiated following its departure from the JCPOA (News, 2018).

The increasingly militarized crisis continued into June. On June 6, 2019, Iranian-backed rebels in Yemen shot down a MQ-9 Reaper, leading the US Central Command (CENTCOM) Commander to warn that US forces faced an imminent threat throughout the region (Kube, 2019). On June 13, magnetic mines, likely delivered by Iranian unmanned subsurface vehicles, damaged two additional commercial vessels, leading the United States to announce additional troop deployments.

The downing of a US RQ-4A Global Hawk UAV on June 20, 2019, served notice that conflict was likely to escalate. The United States deemed it an unprovoked attack of an aircraft in international waters. President Trump ordered a military strike on June 20, but halted the operation over fears of mass casualties on the Iranian side, or fears of the impact of a war with Iran on reelection. He stated on Twitter, “We were cocked & loaded to retaliate last night on 3 different sights when I asked, how many will die. 150 people, sir, was the answer from a General. 10 minutes before the strike I stopped it, not proportionate to shooting down an unmanned drone.” (Olorunnipa et al., 2019).

Instead of escalating the conflict, on June 22 the United States leveraged a series of cyber operations to respond proportionally to Iranian provocations. There seems to have been a few distinct operations; it is unclear how many separate teams or tasks were directed against Iran. One operation disabled Iran’s ability to monitor and track ships in the region by attacking their shipping databases (Barnes, 2019b). Another operation by US Cyber Command was said to have disabled Iranian missile sites, making them vulnerable to air attacks (Nakashima, 2019). In addition, the United States was also likely dumping Iranian code on the site VirusTotal (Vavra, 2019), potentially impairing Iranian’s ability to retaliate by spilling their tools so other defenders were prepared.

The cyber operations served to signal risk to the Iranians and preserve further options to manage the crisis if it was to continue. The proportional response to Iran’s activities possibly allowed for the conflict to stabilize and helped push the two states away from the brink of war. On the road to war, cyber options provide a critical path away from confrontation while still managing to service domestic audience concern

On June 24, cyber security scholar, Bobby Chesney, observed, “Indeed, reading the tea leaves from the past weekend, it appears the cyber option helped ensure there was an off-ramp from a kinetic response that might have led to further escalation.” (Pomerleau & Eversden, 2019). On June 25, Valeriano and Jensen (2019) wrote a column in *The Washington Post* that stated, “contrary to conventional wisdom, cyber options preserve flexibility and provide leaders an off-ramp to war.”

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Following a tense summer, the conflict moved into a new phase in late 2019 and 2020 with the killing of an American contractor after a rocket attack on the US base in Iraq on December 27, 2019 (Barnes, 2019a). The United States retaliated with strikes against Iranian proxies, the Hezbollah, in Iraq and Syria. Hezbollah then attacked the American embassy in Iraq, leading to the US president authorizing the assassination of IRGC Commander, Qasem Solemani, on January 3, 2020 (Zraick, 2020). The United States moved to deploy 4,000 addition troops in the region and Iran retaliated by launching missile strikes on US bases in Iraq, wounding over a hundred soldiers (Zaveri, 2020). The conflict was finally de-escalated, with the United States choosing to not respond to the Iranian attack by claiming that no one had been killed. Since there was six months between the summer and winter 2019/2020 incidents, they are treated as two distinct, albeit linked, crisis cases.

6.4 *Assessing the Case*

Assessment of the events suggests that the crisis with Iran could have escalated in June 2019 after the downing of the Global Hawk UAV, seen as a significant piece of military hardware costing around $220 million (Newman, 2019). Demands for retaliation and escalation were rife in the foreign policy community and within the Trump Administration (Trevithick, 2019).

Instead of escalation, the United States took a different path, consistent with Hypothesis 1. By responding through cyber actions, the United States did two things. First, it demonstrated commitment and credibility to counter Iranian operations by signaling intent for future operations that could have dramatic consequences on Iranian power in the region. Second, these cyber operations also served as Phase 0 operations meant to shape the environment and set the conditions should the United States want to use additional military options in the future. With Iranian defensive systems compromised, Iran was vulnerable to an American attack that never came, and simultaneously subject to a cyber substitute consistent with Hypothesis 3. Cyber operations served to de-escalate the conflict by vividly illustrating the shadow of the future for continued Iranian harassment in the region.

President Trump also increased targeted sanctions directed at Iran's leadership and threated further strikes, stating that he did not need Congressional approval due to the existing authorization for military forces in the region to respond to terrorist threats (Crowley, 2020).[9] These moves are consistent with Hypothesis 2, which suggests that cyber operations are used to complement other forms of power if there is a consideration for escalation.

[9] A list of all US sanctions can be found at a US State Department resource (www.state.gov/iran-sanctions/). Sanctions were already fairly extensive in the summer of 2019 and the United States only added targeted sanctions against industries and various actors after the downing of the US Global Hawk.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

When challenged by a strike on an American asset in the region, the United States had two options, respond in kind or escalate the conflict. Doing nothing would incur significant audience costs among President Trump's base of support because it would demonstrate weakness. Escalation would likely provoke retaliation by proxy forces all over the Middle East leading to significant US casualties. War would also harm the President's reelection chances after promising a reduction in tensions and an end to the wars in the region (Tesler, 2020).

Choosing the option of cyber operations and increased sanctions fits clearly with an off-ramp perspective on crisis bargaining. As Hypothesis 3 argued, cyber operations are likely to be used as substitutes when there are no indications of adversary cyber activity. Here cyber options substituted military options because Iran did not escalate in the cyber domain in response to US cyber moves, and Washington likely judged it had a domain advantage.

Cyber options offered a path out of the conflict through responding in ways that target Iran's command and control functions directly, demonstrating decreased capacity for Iran to control their battlespace. Of particular interest, some of the cyber operations specifically limited Iran's ability to retaliate in cyberspace by leaking the malicious code Tehran was likely to use. No other military response options were utilized, although they were considered, after cyber operations were leveraged. Cyber options can serve as off-ramps from the path to war.

7 CONCLUSION: THE PROMISE AND LIMIT OF CYBER OFF-RAMPS

Based on the observations from experiments and a case study of a US-Iranian crisis in the summer of 2019, we conclude that cyber response options limit the danger of escalation. If used correctly to signal to the opposition to moderate behavior, or as demonstrations of resolve, cyber operations allow states to check the behavior of the opposition with minimal danger of escalation. Cyber options allow a state to express discontent and reshape the balance of information between two opposing parties.

To date, states appear to use cyber options to decrease tensions. This is a counterintuitive finding when many in the discipline suggest that either cyber is inherently escalatory or the nature of conflict has changed. It might be true that conflict has changed, but information operations and cyber operations are generally less escalatory and therefore less dangerous than confronting the opposition with conventional weapons. In other words, the logic of substitution and complements appears to apply to the digital domain. The nature of research suggests that there is less danger in using cyber operations as off-ramps to initial confrontations. We must be clear that we are not suggesting cyber operations as a first strike option. To the contrary, cyber operations likely risk sparking a security dilemma when the target is less capable. Yet, as reactions to initial hostility, cyber options provide a path away from war.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Despite a demonstrated case, as well empirical and experimental evidence suggesting cyber operations are not associated with crisis escalation, there are still limits to these findings. Inequality and the inability of a state to respond to a cyber action with cyber response options increases the dangers of escalation. The behavior and strategic posture of the target can be a critical part of the equation. A history of disputes that create overall tension in a dyad can lead to escalation if the issue is salient enough, even if there are cyber response options (Vasquez, 1993). Our simulation was constricted to one interaction, meaning that we did not test the conditions for escalation across a series of disputes.

The policy advice that emerges from this research is to integrate cyber options into a "whole of government" response tailored to each contingency. In an extended bargaining situation, cyber responses to initial moves can reveal information and decrease tensions, countering much of the hype and hysteria about digital technology exacerbating conflict. That said, cyber operations must be evaluated in terms of the extent to which they act as a complement or substitute, as well as how they might lead to misperception or undermine global connectivity, given the fact that the networks cyber operations target and rely on are largely owned by the private sector. Misperception is still a risk in the digital domain.

The policy goal should be to adopt moderate cyber operations that seek to shape the environment to avoid escalation risks, even if those risks are generally low. By revealing and gathering information in a bargaining situation, cyber options can help decrease tensions by giving states the space they need to maneuver and seek to end a conflict. Using cyber operations, especially cyber operations meant to critically wound command and control facilities or cause death in an offensive manner early during the precrisis period, would likely lead to escalation.

REFERENCES

Al-Aloosy, M. (2020). *The changing ideology of Hezbollah*. Springer.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489), 1390–1396.

Axelrod, R., & Keohane, R. O. (1985). Achieving cooperation under anarchy: Strategies and institutions. *World Politics*, 38(1), 226–254.

Barnes, J. E. (2019a, December 27). American contractor killed in rocket attack in Iraq. *New York Times*. www.nytimes.com/2019/12/27/us/politics/american-rocket-attack-iraq.html

Barnes, J. E. (2019b, August 28). U.S. cyberattack hurt Iran's ability to target oil tankers, officials say. *New York Times*. www.nytimes.com/2019/08/28/us/politics/us-iran-cyber-attack.html

Barzashka, I. (2013). Are cyber-weapons effective? Assessing Stuxnet's impact on the Iranian enrichment programme. *The RUSI Journal*, 158(2), 48–56.

Beardsley, K., & Asal, V. (2009a). Nuclear weapons as shields. *Conflict Management and Peace Science*, 26(3), 235–255.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Beardsley, K., & Asal, V. (2009b). Winning with the bomb. *Journal of Conflict Resolution*, 53(2), 278–301.

Bolton, J. (2019, May 5). *Statement from the National Security Advisor Ambassador John Bolton*. White House. www.whitehouse.gov/briefings-statements/statement-national-security-advisor-ambassador-john-bolton-2/

Booth, K., & Wheeler, N. (2007). *The security dilemma: Fear, cooperation, and trust in world politics*. Springer Nature.

Borghard, E. D., & Lonergan, S. W. (2017). The logic of coercion in cyberspace. *Security Studies*, *26*(3), 452–481.

Braithwaite, A., & Lemke, D. (2011). Unpacking escalation. *Conflict Management and Peace Science*, 28(2), 111–123.

Buchanan, B. (2016). *The cybersecurity dilemma: Hacking, trust, and fear between nations*. Oxford University Press.

Carson, A. (2020). *Secret wars: Covert conflict in international politics*. Princeton University Press.

Clarke, R. A., & Knake, R. K. (2014). Cyber war. Old Saybrook: Tantor Media, Incorporated.

Craig, A., & Valeriano, B. (2016). *Conceptualising cyber arms races* [Manuscript]. 8th International Conference on Cyber Conflict Tallinn, Estonia.

Crowley, M. (2020, May 6). Trump vetoes measure demanding congressional approval for Iran conflict. *New York Times*. www.nytimes.com/2020/05/06/us/politics/trump-vetoes-iran-war-powers.html

Dunning, T. (2016). Transparency, replication, and cumulative learning: What experiments alone cannot achieve. *Annual Review of Political Science*, 19(1), 541–563.

Eckstein, H. (1975). Case studies and theory in political science. In F. Greenstein & N. Polsby (Eds.), *Handbook of political science* (vol. 7, pp. 79–138). Reading, MA: Addison-Wesley.

Eisenstadt, M. (2016). *Iran's lengthening cyber shadow*. Washington Institute for Near East Policy.

Eisenstadt, M. (2017). *Iran after sanctions: Military procurement and force-structure decisions*. International Institute for Strategic Studies. www.washingtoninstitute.org/uploads/Documents/opeds/Eisenstadt20171219-IISS-chapter.pdf

Fearon, J. D. (1995). Rationalist explanations for war. *International Organization*, 49(3), 379–414.

Fitzpatrick, M. (2017). Assessing the JCPOA. *Adelphi Series*, 57(466–467), 19–60.

Gartzke, E., & Lindsay, J. R. (2019). *Cross-domain deterrence: Strategy in an era of complexity*. Oxford University Press.

Glaser, C. L. (1997). The security dilemma revisited. *World Politics*, 50(1), 171–201.

Healey, J., & Grindal, K. (2013). *A fierce domain: Conflict in cyberspace, 1986 to 2012*. Cyber Conflict Studies Association.

Hensel, P. R., & Mitchell, S. M. (2017). From territorial claims to identity claims: The Issue Correlates of War (ICOW) Project. *Conflict Management and Peace Science*, 34(2), 126–140.

Herz, J. H. (1950). Idealist internationalism and the security dilemma. *World Politics*, *2*(2), 157–180.

Hubbard, B., Karasz, P., & Reed, S. (2019, September 14). Two major Saudi oil installations hit by drone strike, and U.S. blames Iran. *New York Times*. www.nytimes.com/2019/09/14/world/middleeast/saudi-arabia-refineries-drone-attack.html

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2016). More similar but less satisfying: Comparing preferences for and the efficacy of within-and cross-category substitutes for food. *Psychological Science*, 27(6), 894–903.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Huth, P. K. (1999). Deterrence and international conflict: Empirical findings and theoretical debates. *Annual Review of Political Science*, *2*(1), 25–48.

Hyde, S. D. (2015). Experiments in international relations: Lab, survey, and field. *Annual Review of Political Science*, *18*(1), 403–424.

Jensen, B. (2017). The cyber character of political warfare. *The Brown Journal of World Affairs*, *24*(1), 159.

Jensen, B., & Valeriano, B. (2019a, March 27). *Cyber escalation dynamics: Results from war game experiments international studies association*. Annual Meeting Panel: War Gaming and Simulations in International Conflict.

Jensen, B., & Valeriano, B. (2019b). *What do we know about cyber escalation? Observations from simulations and surveys*. Atlantic Council. www.atlanticcouncil.org/wp-content/uploads/2019/11/What_do_we_know_about_cyber_escalation_.pdf

Jensen, B., & Work, J. D. (2018, September 4). *Cyber civil-military relations: Balancing interests on the digital frontier*. War on the Rocks. https://warontherocks.com/2018/09/cyber-civil-military-relations-balancing-interests-on-the-digital-frontier/

Jervis, R. (1978). Cooperation under the security dilemma. *World Politics*, *30*(2), 167–214.

Jervis, R. (2017). *Perception and misperception in international politics*. Princeton University Press.

Johnston, T., Lane, M., Casey, A., Williams, H. J., Rhoades, A. L., Sladden, J., Vest, N., Reimer, J. R., & Haberman, R. (2020). *Could the Houthis be the next Hizballah? Iranian proxy development in Yemen and the future of the Houthi movement*. RAND Corporation.

Kaplan, F. (2016). *Dark territory: The secret history of cyber war*. Simon & Schuster.

Kello, L. (2017). *The virtual weapon and international order*. Yale University Press.

Kinzer, S. (2008). *All the Shah's men: An American coup and the roots of Middle East terror*. John Wiley & Sons.

Kirkpatrick, D. D., Perez-Pena, R., & Reed, S. (2019, June 13). Tanks are attacked in the Mideast, and U.S. says video shows Iran war involved. *New York Times*. www.nytimes.com/2019/06/13/world/middleeast/oil-tanker-attack-gulf-oman.html

Kostyuk, N., & Zhukov, Y. M. (2019). Invisible digital front: Can cyber attacks shape battlefield events? *Journal of Conflict Resolution*, *63*(2), 317–347.

Kreps, S., & Schneider, J. (2019). Escalation firebreaks in the cyber, conventional, and nuclear domains: Moving beyond effects-based logics. *Journal of Cybersecurity*, *5*(1), 1–11. https://doi.org/10.1093/cybsec/tyz007

Krugman, P., & Wells, R. (2008). *Microeconomics*. Macmillan.

Krugman, P. R., Robin, W., & Olney, M. L. (2008). *Fundamentals of economics*. Reversed.

Kube, C. (2019, June 6). U.S. Commander says American Forces face "Imminent" threat from Iran. *NBC News*. www.nbcnews.com/news/military/u-s-commander-says-american-forces-face-imminent-threat-iran-n1014556

Lerner, K. L. (2020). *The American Assassination of Iranian Gen. Qassem Soleimani: Strategic Implications, Asymmetrical Threat Risks, and US Congressional Reporting Requirements*. Taking Bearings.

Levy, J. S. (2008). Case studies: Types, designs, and logics of inference. *Conflict Management and Peace Science*, *25*(1), 1–18.

Libicki, M. C. (2012). *Crisis and escalation in cyberspace*. RAND Corporation.

Lin-Greenberg, E., Pauly, R., & Schneider, J. (2020, August 18). *Wargaming for political science research*. SSRN. http://dx.doi.org/10.2139/ssrn.3676665

Lindsay, J. R. (2013). Stuxnet and the limits of cyber warfare. *Security Studies*, *22*(3), 365–404.

Lindsay, J. R., & Gartzke, E. (2018). Coercion through cyberspace: The stability-instability paradox revisited. In K. M. Greenhill & P. Krause (Eds.), *Coercion: The power to hurt* (pp. 179–203). Oxford University Press.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Maness, R., Valeriano, B., & Jensen, B. (2019). *Dyadic cyber incident and campaign dataset* (Version 1.5) [Data File].

Marshall, A. (1890). The principles of economics. McMaster University Archive for the History of Economic Thought.

Martelle, M. (2018, August 13). *Joint Task Force ARES and Operation* GLOWING SYMPHONY: *Cyber Command's Internet War against ISIL*. National Security Archive. https://nsarchive.gwu.edu/briefing-book/cyber-vault/2018-08-13/joint-task-force-ares-operation-glowing-symphony-cyber-commands-internet-war-against-isil

Milner, H. V., & Tingley, D. H. (2011). Who supports global economic engagement? The sources of preferences in American foreign economic policy. *International Organization*, 65(1), 37–68.

Most, B. A., & Starr, H. (1983). International relations theory, foreign policy substitutability, and nice laws. *World Politics*, 36(3), 383–406.

Most, B. A., & Starr, H. (2015). *Inquiry, logic, and international politics: With a new preface by Harvey Starr*. University of South Carolina Press.

Mousavian, S. H., & Toossi, S. (2017). Assessing US–Iran nuclear engagement. *The Washington Quarterly*, 40(3), 65–95.

Nakashima, E. (2019, June 22). Trump approved cyber-strikes against Iran's missile systems. *The Washington Post*. www.washingtonpost.com/world/national-security/with-trumps-approval-pentagon-launched-cyber-strikes-against-iran/2019/06/22/250d3740-950d-11e9-b570-6416efdc0803_story.html

Nasri, F. (1983). Iranian studies and the Iranian Revolution. *World Politics*, 35(4), 607–630.

Newman, L. H. (2019, June 20). The drone Iran shot down was a $220M surveillance monster. *Wired*. www.wired.com/story/iran-global-hawk-drone-surveillance/

News, B. (2018, August 7). Iran sanctions: Trump warns trading partners. *BBC News*. www.bbc.com/news/world-us-canada-45098031

Olorunnipa, T., Dawsey, J., Demirjian, K., & Lamothe, D. (2019, June 21). "I stopped it": Inside Trump's last-minute reversal on striking Iran. *The Washington Post*. www.washingtonpost.com/politics/i-stopped-it-inside-trumps-last-minute-reversal-on-striking-iran/2019/06/21/e016effe-9431-11e9-b570-6416efdc0803_story.html

Palmer, G., & Bhandari, A. (2000). The investigation of substitutability in foreign policy. *Journal of Conflict Resolution*, 44(1), 3–10.

Pauly, R. B. (2018). Would US leaders push the button? Wargames and the sources of nuclear restraint. *International Security*, 43(2), 151–192.

Perla, P. P. (1990). *The art of wargaming: A guide for professionals and hobbyists*. Naval Institute Press.

Pomerleau, M., & Eversden, A. (2019, June 24). *What to make of US cyber activities in Iran*. Fifth Domain. www.fifthdomain.com/dod/2019/06/25/why-trump-may-have-opted-for-a-cyberattack-in-iran/

Powell, R. (2002). Bargaining theory and international conflict. *Annual Review of Political Science*, 5(1), 1–30.

Pytlak, A., & Mitchell, G. E. (2016). Power, rivalry and cyber conflict: An empirical analysis. In K. Fris & J. Ringsmose (Eds.), *Conflict in cyber space: Theoretical, strategic and legal perspectives* (pp. 81–98). Routledge.

Ramazani, R. K. (1989). Iran's foreign policy: Contending orientations. *Middle East Journal*, 43(2), 202–217.

Reddie, A. W., Goldblum, B. L., Lakkaraju, K., Reinhardt, J., Nacht, M., & Epifanovskaya, L. (2018). Next-generation wargames. *Science*, 362(6421), 1362–1364.

Renshon, J. (2015). Losing face and sinking costs: Experimental evidence on the judgment of political and military leaders. *International Organization*, 69(3), 659–695.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Reynolds, N. (2019). *Putin's Not-so-secret Mercenaries: Patronage, geopolitics, and the Wagner group*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2019/07/08/putin-s-not-so-secret-mercenaries-patronage-geopolitics-and-wagner-group-pub-79442

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.

Roff, H. (2016, September 28). "Weapons autonomy risk is rocketing." Foreign Policy. https://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/

Rovner, J. (2019, September 16). *Cyber war as an intelligence contest*. War on the Rocks. https://warontherocks.com/2019/09/cyber-war-as-an-intelligence-contest/

Sample, S. G. (1997). Arms races and dispute escalation: Resolving the debate. *Journal of Peace Research*, *34*(1), 7–22.

Schelling, T. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.

Schelling, T. C. (1958). The strategy of conflict. Prospectus for a reorientation of game theory. *Journal of Conflict Resolution*, *2*(3), 203–264.

Schelling, T. C. (1966). *Arms and influence*. New Haven: Yale University Press.

Schelling, T. C. (2020). *Arms and influence*. Yale University Press.

Schmitt, E., & Barnes, J. E. (2019, May 13). White House reviews military plans against Iran, in echoes of Iraq war. *New York Times*. www.nytimes.com/2019/05/13/world/middleeast/us-military-plans-iran.html

Schneider, J. (2017). *Cyber and crisis escalation: Insights from Wargaming*. USASOC Futures Forum. https://paxsims.files.wordpress.com/2017/01/paper-cyber-and-crisis-escalation-insights-from-wargaming-schneider.pdf

Schneider, J. (2019). The capability/vulnerability paradox and military revolutions: Implications for computing, cyber, and the onset of war. *Journal of Strategic Studies*, *42*(6), 841–863.

Sechser, T. S., & Fuhrmann, M. (2017). *Nuclear weapons and coercive diplomacy*. Cambridge University Press.

Shay, S. (2017). *The axis of evil: Iran, Hizballah, and the Palestinian Terror*. Routledge.

Sheskin, D. J. (2020). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall.

Simon, S. (2018). Iran and President Trump: What is the endgame? *Survival*, *60*(4), 7–20.

Slayton, R. (2017). What is the cyber offense-defense balance? Conceptions, causes, and assessment. *International Security*, *41*(3), 72–109.

Sniderman, P. M. (2018). Some advances in the design of survey experiments. *Annual Review of Political Science*, *21*(1), 259–275.

Starr, H. (2000). Substitutability in foreign policy: Theoretically central, empirically elusive. *Journal of Conflict Resolution*, *44*(1), 128–138.

Straub, J. (2019). Mutual assured destruction in information, influence and cyber warfare: Comparing, contrasting and combining relevant scenarios. *Technology in Society*, 59, 101177.

Tabatabai, A. M. (2020). After Soleimani: What's next for Iran's Quds force? *CTC Sentinel*, *13*(1), 28–33.

Tesler, M. (2020, January 4). Attacking Iran will not help Trump win reelection. Here's why. *The Washington Post*. www.washingtonpost.com/politics/2020/01/04/attacking-iran-wont-help-trump-win-reelection-heres-why/

Thompson, W., & Dreyer, D. (2011). *Handbook of international rivalries*. CQ Press.

Toft, M. D. (2014). Territory and war. *Journal of Peace Research*, *51*(2), 185–198.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Trevithick, J. (2019, June 20). *No easy decisions for U.S. over how to react to Iran shooting down navy drone*. The Drive. www.thedrive.com/the-war-zone/28626/no-easy-decisions-for-u-s-over-how-to-react-to-iran-shooting-down-navy-drone

Trump, D. (2018, May 8). *President Donald J. Trump is ending United States participation in an unacceptable Iran deal*. White House. www.whitehouse.gov/briefings-statements/president-donald-j-trump-ending-united-states-participation-unacceptable-iran-deal/

Valeriano, B. (2013). *Becoming rivals: The process of interstate rivalry development*. Routledge.

Valeriano, B., & Jensen, B. (2019, January 15). *The myth of the cyber offense: The case for cyber restraint*. Cato Institute. www.cato.org/publications/policy-analysis/myth-cyber-offense-case-restraint

Valeriano, B., & Jensen, B. (2019, June 25). How cyber operations can help manage crisis escalation with Iran. *The Washington Post*. www.washingtonpost.com/politics/2019/06/25/how-cyber-operations-can-help-manage-crisis-escalation-with-iran/

Valeriano, B., Jensen, B. M., & Maness, R. C. (2018). *Cyber strategy: The evolving character of power and coercion*. Oxford University Press.

Valeriano, B., & Maness, R. C. (2014). The dynamics of cyber conflict between rival antagonists, 2001–11. *Journal of Peace Research*, 51(3), 347–360.

Valeriano, B., & Maness, R. C. (2015, May 13). The coming cyberspace: The normative argument against cyberwarfare. *Foreign Affairs*. www.foreignaffairs.com/articles/2015-05-13/coming-cyberpeace

Valeriano, B., & Maness, R. C. (2015). *Cyber war versus cyber realities: Cyber conflict in the international system*. Oxford University Press.

Van Creveld, M. (2013). *Wargames: From gladiators to gigabytes*. Cambridge University Press.

Vavra, S. (2019, July 10). *Why cyber command's latest warning is a win for the government's information sharing efforts*. CyberScoop. www.cyberscoop.com/cyber-command-information-sharing-virustotal-iran-russia/

Vasquez, J. A. (1993). *The war puzzle*. Cambridge University Press.

Vasquez, J. A., & Henehan, M. T. (2010). *Territory, war, and peace*. Routledge.

Wise, H. (2013). *Inside the danger zone: The US Military in the Persian Gulf, 1987–1988*. Naval Institute Press.

Wong, E., & Schmitt, E. (2019, April 8). Trump designates Iran's revolutionary guards a foreign terrorist group. *New York Times*. www.nytimes.com/2019/04/08/world/middleeast/trump-iran-revolutionary-guard-corps.html

Yee, V. (2019, May 13). Claim of attacks on 4 oil vessels raises tensions in the Middle East. *New York Times*. www.nytimes.com/2019/05/13/world/middleeast/saudi-arabia-oil-tanker-sabotage.html

Zaveri, M. (2020, February 10). More than 100 troops have brain injuries from Iran missile strike, Pentagon says. *New York Times*. www.nytimes.com/2020/02/10/world/middleeast/iraq-iran-brain-injuries.html

Zraick, K. (2020, January 3). What to know about the death of Iranian General Suleimani. *New York Times*. www.nytimes.com/2020/01/03/world/middleeast/suleimani-dead.html

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

5

Cyber Peace and Intrastate Armed Conflicts

Toward Cyber Peacebuilding?

Jean-Marie Chenou and John K. Bonilla-Aranzales

1 INTRODUCTION

South Africa is a renowned case for its remarkable peacebuilding process that followed the transition from the apartheid era in the 1990s, particularly in terms of reconciliation, restorative justice, forgiveness, and healing from a violent past (Borris, 2002). However, the reconciliation process is ongoing, as seen in the first five days of September 2019 when some xenophobic, looting, and violent attacks emerged in Johannesburg. This time, the victims of those violent attacks were not black South Africans. Instead, the victims were Nigerians who lived and worked in South Africa (Holmes, 2019). This episode of violence could be impacted by different factors, including social media promotion. This example highlights a common feature of online communication in conflict-torn and postconflict societies in various parts of the world. The digital transformation has blurred the boundaries between cyberspace and "physical" space, creating a continuum between online and offline violence. As such, cyberspace has become a realm for political confrontation. Information and data can both be tools to empower dissidents while also being weapons for users, decision makers, governments, and armed groups (Berman, Felter & Shapiro, 2020; Duncombe, 2019). In this context, threats of violence are published on webpages and social media platforms to create and exacerbate a climate of fear. Violence targeted at specific minority groups reproduces offline practices of discrimination and hatred (Alexandra, 2018). Moreover, social media and messaging applications are used to mobilize populations generating large-scale collective actions that have created meaningful changes or call for actions worldwide, such as the cases of the Arab Spring (Salem, 2014), the Black Lives Matter movement in the United States (Zeitzoff, 2017), or the feminist movement in Argentina (Chenou and Másmela, 2019). These dynamics are particularly important in postconflict contexts where new opportunities for truth and reconciliation emerge while conflictual relationship might migrate online.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Many cybersecurity studies focus on state actors and, more specifically, on great powers with strong capacities to conduct cyber operations on a global scale, such as the Stuxnet attack (Valeriano and Maness, 2018), or the digital attack on the Ukrainian power grid in 2015 (Deibert 2018). However, the resolution of intrastate conflict dynamics, which are crucial elements undermining the existence of a sustainable, stable, and secure cyberspace, usually goes ignored. The use and impact of Information and Communication Technologies (ICTs) in cyberspace during intrastate conflicts has also drawn much attention due to its impact, expanding the analysis of the media's role in conflicts. However, cyberspace's role in peacebuilding has been less studied, despite the Tunis Commitment for the Information Society, adopted by the UN in 2005, which acknowledges the potential of ICTs to promote peace by "assisting post-conflict peacebuilding and reconstruction" (United Nations, 2005). As illustrated by the aforementioned South African riots case, the issue of peacebuilding in cyberspace goes beyond access and safe use of technology. It also includes the regulation of violent content and information. This chapter proposes a dialogue between Internet studies and the analysis of peacebuilding to define the notion of cyber-peacebuilding based on the cases of Colombia and South Africa. Drawing upon the four pillars of cyber peace (Shackelford, 2020, preface), it identifies the main venues for cyber peacebuilding research. We propose a working definition of cyber peacebuilding as those activities that delegitimize online violence, build capacity within society to peacefully manage online communication, and reduce vulnerability to triggers that may spark online violence. These efforts include, but are not limited to, the prevention of the use of online violence as a conflict reduction strategy. They also seek to address the structural causes of conflict by eliminating online discrimination, detecting possible threats and power abuses, and promoting inclusion and peaceful communication in cyberspace.

This chapter, organized into three parts, contributes to structuring the emerging field of cyber peacebuilding research. It draws a bridge between cyber peace, understood as a global public good, and its implementation at the national level by drawing on the cases of South Africa and Colombia.

It begins by broadening the perspective of cyber peace studies to include intrastate armed conflicts located mostly in the Global South. The second section outlines the challenges posed by intrastate conflicts for global cyber peace and draws upon cybersecurity and conflict resolution literature to define cyber-peacebuilding. The third section focuses on how the four pillars of cyber peace used as a framework in this volume – namely human rights, access and cybersecurity norms, multistakeholder governance, and stability – can help structure cyber peacebuilding research and even inform policymakers with a particular focus on South Africa and Colombia. Finally, the chapter concludes with the relevance of cyber peacebuilding research and draws some examples for further research on the issue.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

2 TOWARD A COMPREHENSIVE CYBER PEACEBUILDING APPROACH

The use of ICTs both affects the dynamics of violent disputes and helps to generate peacebuilding activities (Puig, 2019). The use of these technologies does not follow a deterministic path. Technologies, including social media platforms, provide new ways of communication between parties that could increase harm as well as provide novel forms of cooperation. To better understand their impact, we explore some challenges that intrastate armed conflicts generate in a global scenario, then we discuss the role of cyberspace in internal conflicts. Finally, we propose some ideas about the relevance of cyber peacebuilding based on intrastate conflict resolution scenarios.

Intrastate armed conflicts have emerged as a new complex challenge globally, particularly in the Global South (Pettersson & Öberg, 2020). A substantial increase in intrastate disputes occurred in the post–Cold War period, becoming the most frequent and deadly form of armed conflict in the world (Mason & Mitchell, 2016), with devastating consequences at social and psychological levels (Wallensteen, 2018). Intrastate armed conflict can be defined as civil wars (Sarkees & Wayman, 2010) or understood as asymmetric conflicts (Berman, Felter & Shapiro, 2020). Intrastate armed conflicts include periods of military hostility between government security forces and members of one or more armed opposition groups within a state lasting ten or more days, without regard to the number of fatalities (Mullenbach, 2005). They can be categorized according to the dispute's issue and the rebels' goals, such as ideological revolutions, ethnic revolutions, and secessionist revolts. Moreover, they can be characterized by the causes of their occurrences. Internal armed conflicts can be explained by *greed*, centered on individuals' desire to maximize their profits; *grievance*, where conflict occurs as a response to socioeconomic or political injustice; and *opportunity*, which highlights factors that make it easier to engage in violent mobilizations (Cederman & Vogt, 2017).

The role of cyberspace in internal conflicts can be interpreted as a double-edged sword, as it enhances the interaction between users, digital platforms, and governmental agencies across multiple technological devices. However, the tensions concerning its positive or negative use not only depend on the users, who range from ordinary citizens to political leaders, rebels, and extremist groups, among other societal actors – all of whom interact using ICTs. The social and political contexts of its use are relevant because those conditions allow for the presence of new actors that behave with complex rules, which undoubtedly change the dynamics of civil wars and peacebuilding scenarios. In short, cyberspace matters in the development and ending of intrastate conflicts because they have become information centric (Berman et al., 2020; Steinberg, Loyle, & Carugati, in this volume).

Cyberspace capabilities contribute to the creation and tracking of analytical elements concerning the tensions, positions, narratives, and changes in the domestic balance of power of states and non-state actors. It offers the possibility to develop conflict prevention actions, as discussed in Chapter 4. Moreover, cyberspace represents a nurturing ground that allows for the generation and promotion of conflict

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

resolution initiatives. As Ramsbotham, Miall, and Woodhouse (2016, p. 432) argued, the "virtual world of cyberspace is, therefore, contested and conflictual in the same way as the 'real' world is, but the challenges are the same in the sense that emancipatory agendas of conflict resolution apply as much to cyber peacemaking as to 'conventional' peacemaking." In short, this digital space represents a hybrid and dynamic environment (Gohdes, 2018), in which uncertainty and threats emerge, but also where the conflicting parties can create peaceful ways to coexist.

The potential for utilizing cyberspace in peacebuilding activities, particularly to enhance the role of mediators and generate policy change, is a positive example of such technologies (Tellidis & Kappler, 2016; Puig Larrauri & Kahl, 2013). A relatively recent development in cyberspace is the emergence of social media, where users can create content and interact across both micro and macro communities (Kaplan & Haenlein, 2012). The use of social media has undoubtedly changed how we communicate and relate to our world. Its negative uses have raised new complex concerns about ethical and security issues. The recruitment of extremists (Weimann, 2016; Walter, 2017), the increasing polarization among the minority groups who are most active in discussions about public affairs (Barberá, 2020), and the promotion of hate speech (Mathew et al., 2019) are some negative uses that heighten conflict dynamics, not only in cyberspace but also in physical space. However, the use of social media also reduces the costs of information distribution in the framework of violent conflict (Hochwald, 2013), which could generate new social mobilizations and reduce collective action problems (Margetts et al., 2015). Additionally, social media can generate new data and information about the conflict environment that might forecast new violent actions. Its use is also a critical factor in the promotion of narratives that could establish peaceful engagement using a bottom-up approach and could even help foster polycentric information sharing, as was discussed in Chapter 3.

Given the background, our definition of cyber peacebuilding draws upon different strands of literature. Previous efforts to analyze the role of ICTs in the termination of conflicts include cyber peacekeeping and the ICTs for peace frameworks. Moreover, we subscribe to the positive definition of peace adopted by cyber peace scholars. Finally, our definition of cyber peacebuilding is based on a contemporary conflict resolution approach that echoes critical cybersecurity perspectives.

Along with the diffusion of interactions into cyberspace in conflict-torn and postconflict countries, the role of ICTs in peacekeeping operations, and as tools to promote peace, has been increasingly acknowledged by scholars and intergovernmental organizations. From the use of big data in peacekeeping operations (Karlsrud, 2014) to the institutionalization of cyber peacekeeping teams and operations in the United Nations, such as the United Nations' Digital Blue Helmets (Almutawa, 2020; Robinson, et al., 2019; Shackelford, 2020), the literature has broadened to include cyberspace in the analysis of peacekeeping. Cyber peacekeeping is an evolution of an idea that emerged in the 1990s, which posited that ICTs could promote peace. During the process that led to the World Summit on Information Society, the idea of ICTs

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

being used for peace was further developed and included in the Tunis Commitment for the Information Society (United Nations, 2005). However, the use of the concept remained limited in scholarly publications with some exceptions (see Laouris, 2004; Spillane, 2015; Young & Young, 2016) and declined with the massification of social media and the subsequent debate on its role in polarization. While the ICTs for peace scholarship generally focus on access and the infrastructure layer from a techno-optimistic perspective, an analysis of the content layer, and the particular role of social media in conflict and peace dynamics, is a starting point to develop novel inquiries.

Another source of inspiration for cyber peacebuilding is the ongoing effort to promote a positive definition of cyber peace in a scholarly debate primarily dominated by the issue of cyberwar. More specifically, we situate cyber peacebuilding within "the construction of a network of multilevel regimes that promote global, just, and sustainable cybersecurity by clarifying the rules of the road for companies and countries alike to help reduce the threats of cyber conflict, crime, and espionage to levels comparable to other business and national security risks." (Shackelford, 2019, p. 163). In this chapter, we propose an analysis of a cyber peacebuilding approach, which mainly focuses on the national level in postconflict contexts, but also includes the participation of local and international actors. Moreover, the analysis of conflictual contexts and peacebuilding in the digital era can help explore new ways to address the increasing polarization at work in mature democracies.

Finally, cyber peacebuilding adopts a human-centered approach and promotes an emancipatory normative stance on the provision of cybersecurity (Collins, 2020). Within this context, cyber peacebuilding is a reformulation and an extension of the definition of peacebuilding adapted to the digital age. Drawing upon the definition of peacebuilding proposed by the Alliance for Peacebuilding (2012), we define cyber peacebuilding as an active concept that captures those activities that delegitimize online violence, build capacity within society to peacefully manage online communication, and reduce vulnerability to triggers that may spark online violence. Some activities involve, but are not limited to, preventing the use of online violence as a conflict strategy and highlighting the role of users, states, and Big Tech companies in this regard. They also seek to address the structural causes of conflict by eliminating online discrimination; enhancing the scope and impact in the territory of peacebuilding mechanisms; and promoting inclusion and peaceful communication in cyberspace. As such, cyber peacebuilding efforts represent an essential stepping stone in the pursuit of cyber peace as a global public good.

Such a focus on cyber peacebuilding is not entirely new (see, e.g., Puig Larrauri & Kahl, 2013; Tellidis & Kapler, 2016; AlDajani & Muhsen, 2020), even though its expression rarely appears as such. This chapter argues that it can be a useful concept for establishing and structuring a scholarly dialogue that explores the multiple dimensions of peacebuilding in cyberspace beyond a liberal approach, which is often limited to the establishment of liberal institutions – democracy, human rights, open economy, and the rule of law (Zaum, 2012). Here, we adopt a comprehensive approach that includes

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

all activities focused on preventing the causes of violent conflict and strengthen mechanisms to handle conflict in a constructive and nonviolent way (Parlevliet, 2017).

Cyber peacebuilding represents a contribution to global cyber peace from a polycentric approach. Beyond an exclusively top-down perspective on the necessity of global agreements and norms for building a peaceful and stable cyberspace, we adopt a polycentric approach in order to address how local threats to peacebuilding efforts undermine the existence of cyber peace at a global level (for a similar perspective, see Chapter 2). From this perspective, the proliferation of internal armed conflicts requires the construction of peaceful cyber contexts in conflict-torn and postconflict societies.

To further explore the prospects of cyber peacebuilding, we focus on two cases: South Africa and Colombia. With the victory of the African National Congress in the 1994 election, South Africa started a process of transition from the apartheid era, which notably entailed a new constitution and the establishment of a Truth and Reconciliation Commission in 1996. Despite important achievements, the reconciliation process is still ongoing (du Toit, 2017). On the other hand, Colombia has taken a number of major steps toward the termination of a five-decade-long internal conflict. One of the most important was the peace accord of 2016 between the government and the Revolutionary Armed Forces of Colombia (FARC) guerrilla organization. While the two countries are situated at different sites on the conflict/postconflict continuum, they both face the challenges of peacebuilding and reconciliation (Rodríguez-Gómez et al., 2016). Moreover, they are both middle-income and relatively highly digitized countries in the Global South (Chouci and Clark, 2018, p. 163). Also, in both cases, governmental stakeholders have ignored the relevance of cyberspace for the development of peacebuilding actions. Thus, they represent two diverse and interesting cases in which to explore the prospects of cyber peacebuilding.

3 THE FOUR PILLARS OF CYBER PEACEBUILDING

The broad definition of cyber peacebuilding outlined in the previous section encompasses many issues and actors. The four pillars of cyber peace (Shackelford, 2020) provide a framework to structure the analysis. Local threats to cyber peace and cyber peacebuilding efforts can be categorized within the pillars of cyber peace: access and cybersecurity, human rights, multistakeholder governance, and stability (see Figure 5.1).

3.1 *Human Rights, a Call of Action to Update the Social Contract*

The promotion of human rights and peacebuilding mechanisms can be analyzed as joint processes in which peacebuilding insights and methods can advance human rights promotion and protection (Parlevliet, 2017). However, some overlapping tensions must be considered, such as the complicated relationship between freedom of expression and political stability, and the disputes concerning how to handle

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

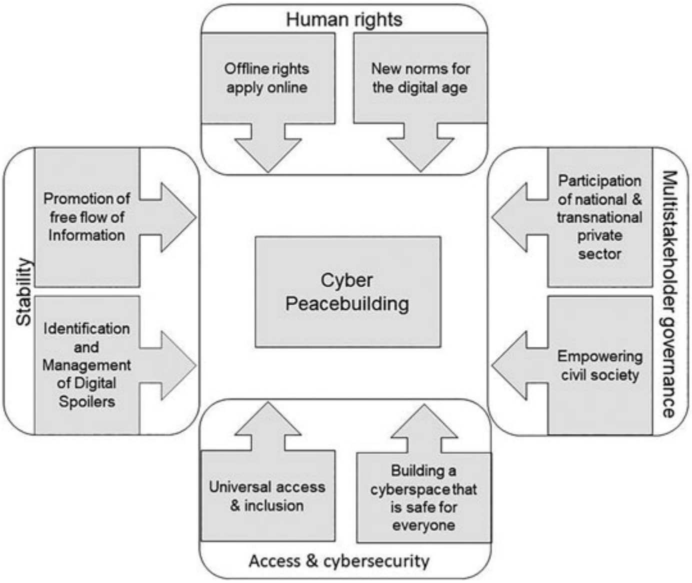

Human rights
Offline rights apply online
New norms for the digital age
Stability
Promotion of free flow of Information
Identification and Management of Digital Spoilers
Cyber Peacebuilding
Multistakeholder governance
Participation of national & transnational private sector
Empowering civil society
Universal access & inclusion
Building a cyberspace that is safe for everyone
Access & cybersecurity

FIGURE 5.1 The contributions of the four pillars of cyber peace to cyber peacebuilding (source: elaborated by the authors [September 21, 2020]).

sensitive issues such hate speech, sexual harassment, and politically driven attacks that foment collective violent responses.

While freedom of expression, privacy, and data protection are covered by International Humanitarian Law and Human Rights Law at the international level (Franklin, 2019; Lubin, 2020), inadequate enforcement mechanisms and profound social issues at the national level, such as a lack of digital literacy and limited Internet access, undermine their implementation (Shackelford, 2019). This difficult adoption of international regulations complicates peacebuilding scenarios because governments regulate freedom of expression to impose an official truth, which sometimes limits the right of expression and association of the opposition sectors. Moreover, in peacebuilding scenarios, some voices, even the official ones, can become radicalized, creating new challenges to stability. In this context, governments can be tempted to prioritize security and stability over freedom of expression and a pluralist dialogue toward peacebuilding.

For example, there is no Internet detailed legal framework in Colombia that guarantees its citizens' fundamental rights in cyberspace. Nevertheless, freedom of speech is viewed comprehensively by the Constitutional Court. It is also backed by Colombia's membership of the Inter-American Human Rights System, which means that this right applies, not only offline, but also in the online world (Dejusticia, Fundación Karisma and Privacy International, 2017). However, the respect of those human rights in cyberspace is often challenged due to the use of a securitization narrative by the current governing party that was an opponent to the peace negotiations

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

with the FARC rebels. The government perceives peacebuilding as a mere process of disarmament, demobilization, and reintegration of former combatants in order to restore stability. This limited view of the peacebuilding process also justifies the use of online state surveillance actions to guarantee national security, in which political leaders, former government officials, journalists, and humanrights activists are targeted because of their support of the peace agreement (Vyas, 2020). Without a doubt, the respect of human rights in Colombia, through cyberspace interactions, represents a new challenge that has been ignored by policymakers in the reconstruction of the social fabric in this transitional society.

Second, there is a tension in cyberspace on how to handle sensitive issues that could evolve into violent conflicts. In this complex scenario, Big Tech companies play a critical role because they are able to track and censor what people post and share. However, in the Global South, this tension is not a priority (Schia, 2018). On the contrary, Big Tech companies are more concerned with access and digitalization than privacy rights. Many social media companies that operate in developing countries do not have clear policies regarding this issue. Instead, their roles in these societies have been linked to increasing disinformation, inciting violence, and decreasing trust in the media and democratic institutions (Bradshaw & Howard, 2019).

The case of South Africa provides an interesting perspective on the respect of human rights in cyberspace as part of a reconciliation process. Their constitution guarantees the right to freedom of opinion and expression. This topic is mainly addressed under the supervision of the South African Human Rights Commission (SAHRC), which was created by the South African Constitution and the Human Rights Commission Act of 1994. Its aim is linked to promoting human rights through a variety of actions about education and raising community awareness; making recommendations to the Parliament; reviewing legislation; and, most importantly, investigating alleged violations of fundamental rights and assisting those affected to secure redress (Sarkin, 1998). Based on its mandate, this institution had provided significant recommendations in the legislation linked to topics data protection (SAHRC, 2012) and recent cybersecurity issues (SAHRC, 2017). Nevertheless, its main challenge is to address issues concerning hate speech and racism in cyberspace, particularly on social media platforms, in a quick and efficient way. This commission acknowledges the issue, and it has taken some steps to face this challenge recognizing the allegations of racism perpetrated on social media (SAHRC, 2016). Most importantly, it started a multistakeholder dialogue to reach a detailed social media charter, including human rights education at all academic levels, to fight racism in the digital sphere (SAHRC, 2019).

In conclusion, in order to address human rights issues in cyberspace, particularly in peacebuilding scenarios, there is a need for a new social contract that recognizes human rights as digital rights. Human rights are considered a crucial element of peacebuilding, which must include cyberspace activities. To provide an impact on the development of peacebuilding mechanisms, some human rights standards, values, and principles must be included. To accomplish that end, some actions

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

concerning public policies regarding security and privacy ought to be addressed by governments without exceeding their power. Big Tech companies must provide stricter and more straightforward privacy protocols and conduct codes in layperson's terms based on the local framework in which they operate. Moreover, civil society's role, particularly that of users, must be present to delimitate the scope of the potential legal actions concerning topics linked to privacy rights, freedom of speech, misinformation, and disinformation. This inclusive approach would help to create a healthy environment for the exchange of ideas and information, enabling all members who coexist in a changing society to respect and resolve their differences, even in the context of intrastate armed conflict and peacebuilding scenarios.

3.2 *Multistakeholder Cyber Peacebuilding*

Multistakeholder governance has become a gold standard in Internet governance and regulations of human activities in cyberspace (Scholte, 2020). While not exempt from criticisms in terms of legitimacy and efficiency, the cooperation between public and private actors has become necessary to handle increasingly large amounts of data and regulate private algorithms and infrastructure, leading to a hybridization of governance (Chenou & Radu, 2019).

This hybridization of governance has also transformed the approach to cybersecurity. Cybersecurity, understood as a national security issue, has historically curtailed the space for multistakeholder governance (Dunn Cavelty, 2013; Kuehn, 2014). However, recent developments in the production and governance of cybersecurity showcase different governance structures beyond the hierarchical state-led governance of cybersecurity (Kuerbis & Badiei, 2017; Mueller, 2017; Shires, 2018; Tanczer et al., 2018). A multi-stakeholder governance of cybersecurity is emerging at the global, national, and local levels (Pernice, 2018). According to Pernice, the shared responsibility in the establishment of cybersecurity and cyber peace requires a:

> […] multilevel and multi-stakeholder system of cybersecurity governance, a system that includes all stakeholders: the individual citizen and civil society, business enterprises, and public authorities, from the local up to the global level (Pernice, 2018, p. 122).

The participation of different sectors in cybersecurity governance is even more important in postconflict contexts, where peacebuilding efforts also require the inclusion of multiple stakeholders (Brzoska et al., 2011; Narten, 2011). Beyond public authorities, three types of actors are of particular importance. First, the private sector plays an essential role in peacebuilding efforts, both during the negotiations and in the implementation of peace agreements (Rettberg, 2007, 2016; Miklian & Schouten, 2019). Second, the media can promote peace and the prevention of incitement to violence (Howard, 2002; Himelfarb & Chabalowski, 2008). Finally, civil society fulfils different functions in peacebuilding, such as: the protection of citizens; the monitoring of human rights violations and the implementation

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

of peace agreements; advocacy for peace and human rights; socialization to values of peace and democracy; intergroup social cohesion; facilitation of dialogue; and service delivery to create entry points for the other functions (Paffenholz, 2010).

Despite some common requirements and goals, multistakeholder cybersecurity governance and multistakeholder peacebuilding are rarely treated together in practice. For example, South Africa has been one of the pioneering countries and a model of multistakeholder peacebuilding with the establishment of an infrastructure for peace. The 1991 National Peace Accord created Regional and Local Peace Committees that were open to any relevant civil society organization, such as religious organizations, trade unions, business and industry representatives, and traditional authorities (Odendaal, 2010). This multistakeholder infrastructure for peace became a reference for further processes (Preventive Action Working Group, 2015). In 1994, South Africa created the National Economic Development and Labour Council in order to allow for multistakeholder participation in the formulation of economic and social policies. However, multistakeholder participation in the governance of cyberspace is limited in South Africa (Mlonzi, 2017). For example, the National Cybersecurity Policy Framework was drafted under the leadership of the South African Department of Communications between 2009 and 2012, but was later transferred to the Ministry of State Security (Global Partners Digital, 2013). As the responsibility of a civilian Ministry, cybersecurity fell under the category of economic and social policy and was thus, open to multistakeholder participation. However, the leadership of the Ministry of State Security limited the scope of cybersecurity and undermined the participation of diverse stakeholders.

In Colombia, multistakeholder participation became institutionalized in economic and social policies through the Consejo Nacional de Política Económica y Social (National Council of Economic and Social Policy). There is a strong participation of diverse stakeholders in the formulation of Internet governance policies organized around the Mesa Colombiana de Gobernanza de Internet (Colombian Internet Governance Forum). Moreover, the recent peace accord acknowledges that "participation and dialogue between different sectors of society contribute to building trust and promoting a culture of tolerance, respect and coexistence" (República de Colombia, 2016, Introducción, translated by the authors). However, the issue of peacebuilding is hardly included in Internet governance debates that tend to reproduce global discussions. On the other hand, the governance of cyberspace is not among the priorities of peacebuilding efforts beyond the question of access (see the section below).

Multistakeholder cyber peacebuilding represents a step further in the implementation of multistakeholder participation. It requires a multistakeholder dialogue between actors involved in the regulation of cyberspace and the diverse sectors that share a responsibility in peacebuilding activities. The cases of South Africa and Colombia illustrate the necessary participation of social media platforms and search engines in peacebuilding efforts. As the corporate social responsibility of digital platforms in campaigns and elections is being discussed in consolidated democracies, the role of digital platforms in postconflict societies to promote peace and

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

limit incitement to violence must be put on the agenda. Likewise, the mass media's responsibility in the promotion of a culture of peace is now shared with new media and social media (Stauffacher et al., 2011; Comninos, 2013). As noted by Majcin (2018), modern peace agreements should include the regulation of social media content that may disrupt the peace and promote the resurgence of violence. These rules could even be institutionalized in the form of special commissions to review content on social media and take action when viral publications undermine peacebuilding.

In sum, multistakeholder governance of cyber peacebuilding entails not only the adoption of national cybersecurity policies that allow for the participation and representation of all stakeholders in postconflict societies, it also requires the adoption of multistakeholder mechanisms directly aimed at the promotion of peace and the prevention of violence in cyberspace with the participation of the private sector, digital platforms, academia, and civil society organizations.

3.3 *Redefining Stability in Cyberspace*

To understand the role of stability in cyberspace, we adopt a nuanced definition of stabilization by drawing upon conflict resolution literature to explain how the tensions generated in cyberspace can affect the dynamics and the conclusion of intrastate conflicts and the development of peacebuilding activities.

There are many approaches to the concept of stability to address armed conflicts. They include issues related to statebuilding (Hoddie & Hartzell 2005), international interventions (Belloni & Moro, 2019), and negotiated peace settlements (Hartzell et al., 2001), among other approaches. From the UN Security Council's vision, stability refers to a desired state of affairs, almost as a synonym of "peace" (Kerttunen & Tikk, 2020). Additionally, this concept has a robust state-centric approach (Carter, 2013). To analyze cyberspace's effect in the ending of intrastate conflicts and peacebuilding scenarios, the dynamic definition proposed by Mielke, Mutschler, and Meininghaus (2020) is more useful. They argue that stability is an open-ended and transformative process which accepts changes in social dynamics to keep its forces in equilibrium by constant reconcilement of interests. In a nutshell, the state's role is crucial to address normative rules, but nonstate actors also play a critical role in achieving long-term stability.

Considering that cyberspace is a very dynamic place, stabilization efforts can lead to the transition from intrastate conflicts toward the restoration of the social fabric through peacebuilding actions. This nuanced approach of stability is crucial to understand issues in conflict resolution scenarios, such as the role of spoilers in cyberspace.

Spoilers can be understood as "key individuals and parties to the armed conflict who use violence or other means to shape or destroy the peace process and in doing so jeopardize the peace efforts" (Nilsson & Söderberg, 2011, p. 624; see also Stedman, 1997). This definition serves to understand the impact of those actors in

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

cyberspace that affect the termination of intrastate conflicts. Digital spoilers are those political actors with relevant influence upon users in cyberspace that exploit their influence to promote violence and spoiling behavior to affect the attempts to achieve peace. They differ from Internet trolls, defined as "unknown online users that create and claim intentionally upsetting statements to enhance strong emotional responses posting offensive or unkind things on the Internet using tactics of disinformation and propaganda" (Petykó, 2018). Digital spoilers are conflicting parties or leaders who use trolling activities, such as the promotion of disinformation and propaganda to affect the achievement of conflict resolution scenarios.

One example of digital spoilers can be found in Colombia, where the opponents of the peace agreement promoted strong and negatively charged hashtags on social media concerning the endorsement of the peace process with the FARC guerilla organization in October 2016 (Nigam et al., 2017). The promotion of these messages, among other factors, affected the perception of the peace negotiations, which was reflected in the rejection of the peace plebiscite by a small margin. The management of spoilers is a daunting task because influential social media platforms users can foment emotions and hostile attitudes against the peacebuilding process. However, these digital spoilers can be tackled when they violate internal regulations of social media platforms (BBC News Mundo, 2019), which highlights the relevance of multistakeholder Internet governance at the national level.

Another relevant example can be found in South Africa, in which political figures use the rhetoric of hate speech toward different communities in order to gain political support (Akhalbey, 2019; Meyer, 2019). The SAHRC has, in the past, analyzed and sanctioned some cases concerning the use of social media to promote hate speech (Geldenhuys and Kelly-Louw, 2020). However, it seems that its mandate does not cover those digital spoilers who express their thoughts in an offensive and disturbing way, pushing the limits of the right to freedom of speech. Their social media statements address critical issues that the peacebuilding process did not solve, such as land reform or race relations, suggesting unpeaceful actions to solve those issues. Additionally, to address the damage that these digital spoilers could make in cyberspace, social media platforms have a key role to play in order to tackle hurtful messages. In this particular case, it seems that there is a misconnection between the conception of the legal rights of freedom of expression provided by the SAHRC and the rules established by social media platforms (Nkanjeni, 2019), which represents a new institutional challenge to address.

In sum, within the framework of cyberspace, stability must be analyzed dynamically. The handling of information plays a critical role because it reflects an age-old tension concerning the relationship between citizens and governments. In that sense, Big Tech companies have become referees and players in a complicated situation. On the one hand, they need to guarantee information and data protection to ensure their legitimacy. On the other hand, they must also respect governmental authority, whose interests are linked to employing surveillance, gathering

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

data, and performing intelligence through controlled information. Amid intrastate armed conflicts and peacebuilding scenarios, the scope of government surveillance could be enhanced, intensifying asymmetric responses. On the other hand, there are more real threats concerning political motivations to spoil conflict resolution scenarios than the risk of cyberspace's misuse of information beyond the cybersecurity framework. Against this background, the concept of digital spoilers is useful to analyze the behavior of actors whose role could substantially affect the dynamics of stability and conflict resolution efforts. This dynamic approach of stability could lead to the fertile ground to develop cyber peacebuilding actions.

3.4 *Inclusion and Human-Centered Cybersecurity*

Universal Internet access is an enabling condition for cyber peace. It was identified as the first of the five principles for cyber peace by the ITU (International Telecommunication Union, 2011). According to the ITU, providing access to telecommunication technologies is part of the responsibilities of states, which was later translated into the (debated) idea of Internet access as a human right (Tully, 2014). However, the relationship between Internet access and cyber peacebuilding is not direct. Access to the Internet is a necessary, though insufficient, condition to building peace that spans offline and online spaces.

Contrary to the late twentieth century's techno-optimistic visions, the "old" concept of the digital divide remains relevant today (van Dijk, 2020). While early accounts of the digital divide focused on physical access and the divide among countries, contemporary analysis of the digital divide insists on the quality of access and the importance of the gap between Internet access within the same country. This dimension is of utmost importance for cyber peacebuilding (Wilson & Wilson, 2009). Those communities that do not have access to the Internet are generally communities that have been historically marginalized (Tewathia et al., 2020). The digital divide also presents a gender dimension that undermines women's participation in peacebuilding (Njeru, 2009). Moreover, since telecommunication infrastructures are targets and battlegrounds during conflicts, violence-affected regions are likely to suffer from inadequate or unstable connectivity (Onuoha, 2013; Adeleke, 2020). Furthermore, the national digital divide certainly undermines states' capacities and presence on peripheral territories and, subsequently, their legitimacy (Krampe, 2016). This lack of presence and the complicated access to increasingly digitized public services reinforces the perceived abandonment by the states among marginalized communities.

Both South Africa and Colombia have reached significant rates of access at the national level as a result of economic development and ambitious policies. While just over 50 percent of the world population had access to the Internet at the end of 2019 (International Telecommunication Union, 2020), access rates in South Africa were around 65 percent (DANE, 2020; STATSSA, 2020). However, national digital

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

divides are still important in both countries. For example, over 74 percent of the Gauteng province around Johannesburg and Pretoria benefit from Internet access, compared to just over 46 percent in the poorer province Limpopo, that also has the smallest white South African population in the country (Media Monitoring et al., 2019, p. 12). In Colombia, less than 10 percent of the inhabitants in 700 out of the 1,123 municipalities have Internet access (Quintero & Solano, 2020). These municipalities are located in geographically remote areas that are also the most affected by the internal conflict.

The bridging of the digital divide is related primarily to the telecommunication infrastructure. Another key element is the use of Internet access by individuals and grassroots organizations to participate in the process of peacebuilding through early warnings, grassroots reporting and monitoring, and data collection "from below." Internet access is necessary to engage in political activities, including peacebuilding (Puig Larrauri & Kahl, 2013; Shandler et al., 2019).

While access is a necessary feature to build the conditions for civil society to participate, it is not sufficient to secure meaningful participation. Another crucial condition for cyber peacebuilding is the construction of a cyberspace that is safe for everyone. A broad and emancipatory definition of cybersecurity goes beyond the preservation and defense of critical national infrastructure. It focuses on the general population, both users and nonusers, to build a postconflict cyberspace that is safe for everyone, including former fighters, victims, women, and marginalized communities. However, cybersecurity policies tend to be framed as a response to conflict. For example, research shows that cybersecurity capacity is greater in countries engaged in civil war. However, this capacity seems to aim to crack down on domestic dissent rather than provide secure cyberspace at the national level (Calderaro & Craig, 2020). Even in postconflict contexts, the original state-centered and militarized approach tends to prevail, despite the evolving conditions. As we have seen, the South African National Cybersecurity Policy Framework was first drafted by the Department of Communications. It was later transferred to the Ministry of State Security and finally adopted in 2015 (State Security Agency, 2015). While it briefly mentions "hate speech" and "fundamental rights of South African citizens" (State Security Agency, 2015, pp. 5, 14), the bulk of the document focuses on national security and on the fight against cybercrime. In the same vein, Colombia adopted a Digital Security policy in 2016 that was drafted during the negotiations between the government and the FARC guerrilla organization (CONPES, 2016). However, the document does not mention the postconflict context. It is largely inspired by the OECD discussions on the management of digital risks and thus, focuses on the necessary conditions for the development of trust in Colombian digital markets. On the other hand, the peace accord only mentions ICTs as a way to access public information and public services such as health and education, without acknowledging their role in the peacebuilding process (República de Colombia, 2016).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Contrary to these examples, the institutionalization of cyber peacebuilding should rely on more comprehensive cybersecurity policies that do not reproduce the patterns of great cyber powers to focus on peacebuilding needs in postconflict societies, such as digital literacy and the regulation of hate speech.

4 CONCLUSIONS AND POLICY IMPLICATIONS

South Africa shows us that reconciliation is possible, even in cyberspace. After the violent attacks in Johannesburg mentioned in the chapter introduction, citizens started to promote hashtags and social media campaigns, such as #SayNoToXenophobia, to call for unity, and looking for an end to the violence in this mature peacebuilding scenario (Levitt, 2019). This example also shows us that while cyberspace has undoubtedly affected the dimensions, approaches, and complex dynamics of intrastate conflicts, it can also promote peacebuilding activities to enhance conflict resolution scenarios.

Colombia provides some examples of how transitional justice contributes to cyber peace in terms of Internet access and human rights. Victims and governmental agencies jointly construct the idea of restorative justice through the use of ICTs and digital tools (Chenou, Chaparro-Martínez, & Mora Rubio, 2019). Moreover, this relationship is tested in times of crisis; for example, during the COVID-19 pandemic, where digital tools allow for the continuation of transitional justice (Alfredo Acosta & Zia, 2020). Under certain conditions, the adoption of ICTs by transitional justice tribunals might enhance the efficiency and efficacy of the distribution of justice, allowing both parties to save time by reducing mobilization costs and unnecessary formalities to the minimum. In terms of truth and reconciliation, evidence can be found in the creation of an online news portal that looks to contribute to the reconstruction, preservation, and dissemination of the historical and judicial truth about the Colombian conflict, adopting a bottom-up and in-depth journalism perspective (Verdad Abierta, 2020).

South Africa also provides different examples of cyber peacebuilding. In terms of peaceful social mobilization using ICTs, the use of mobile phones improves organization efficiency, access to information, and strengthens the collective identity of social movements; for example, among members of the Western Cape Anti-Eviction Campaign in 2001 (Chiumbu, 2012). Moreover, in 2015, South African university students protested around the #FeesMustFall hashtag, to demand relevant changes in their education system, such as the decolonization of curricula and a significant increase in government funding for universities (Cini, 2019). But most importantly, with the use of the hashtag #RhodesMustFall, young South Africans provided some analytical elements about how social media could be the way to collectively question the normative memory production to turn the page away from the apartheid era (Bosch, 2017). Despite the criticisms that could be addressed to the SAHRC for the inconsistent sanctioning of hate speech by

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

political leaders, its contribution to the legislative initiatives concerning data protection, and cybersecurity, respectively, is remarkable (SAHRC, 2012, 2017).

In sum, several contributions to the development of peacebuilding activities are fostered by the linkage between the activities of conflict resolution in cyberspace and in the physical world. This chapter proposed a working definition of cyber peacebuilding in order to provide a broad perspective that reflects changes in the way cyberspace is perceived during interstate armed conflicts and afterwards. ICTs are not only tools, they also constitute and enable the interactions that comprise the lifeblood of cyberspace, transforming the political dynamics of conflict and peacebuilding. Hence, this approach responds to the necessity to implement peacebuilding efforts both in the physical space and in cyberspace. The construction of a stable and lasting peace after intrastate conflicts requires delegitimizing online violence, capacity building within society toward peaceful online communication, and a reduction of the vulnerability to digital spoilers. The structural causes of conflict must also be addressed by eliminating online discrimination and by promoting inclusion and peaceful communication in cyberspace.

The focus on peacebuilding scenarios points to one of the major sources of instability, both online and offline for many countries in the world. While cybersecurity studies tend to focus on state actors that have important capacities, a human-centered perspective on cybersecurity and cyber peace must address the digital dimension of intrastate conflicts as is discussed further in the essays section by the Cyberpeace Institute.

Most intrastate conflicts take place in the Global South. As the majority of Internet users are now located in the Global South, the combination of ICTs and intrastate conflicts is undermining the efforts toward global cyber peace. However, cyberthreats in the Global South are less visible than in the Global North. The focus on commercial threats and on powerful countries obscures the prevalence of cyberthreats against civil society and in the Global South (Maschmeyer et al., 2020). We argue that the concept of cyber peacebuilding sheds light on the relationship between intrastate conflict and global cyber peace and thus contributes to raising awareness about cyberthreats in the Global South.

The four pillars of cyber peace provide a framework to outline comprehensive cyber peacebuilding efforts. As illustrated by Figure 5.1, they highlight the importance of existing human rights and the necessity to create new norms for the digital age. The pillar of multistakeholder governance sheds light on the role of the private sector, and especially of digital platforms and Big Tech companies, along with civil society, to complement and monitor efforts by states and intergovernmental organizations. Stability in postconflict cyberspace can be implemented through the promotion and preservation of a free flow of information and through the identification and management of digital spoilers that undermine the establishment of peace. Finally, the pillar of access and cybersecurity is particularly important in conflict-prone societies where exclusion and marginalization fuel violence. Moreover,

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

cybersecurity must be understood beyond the implementation by the state of a public policy aimed at the protection of national infrastructure and at the management of digital risk. A human-centered approach is necessary in order to build a cyberspace that is safe for everyone.

This preliminary overview of the different dimensions of cyber peacebuilding in the Colombian and South African cases paves the way for further research on the centrality of cyberspace in the termination of contemporary intrastate conflicts, and for the construction of a stable and lasting peace at a global level. Moreover, it identifies venues for political action. States and international organizations must design new norms of human rights for the digital age along with comprehensive and human-centered cybersecurity policies. Capacity building can empower civil society, foster a safe use of technology, and promote peaceful communication and a culture of peace in cyberspace. Finally, the necessary role of digital platforms must be addressed in order to achieve a meaningful participation and a partnership with states and intergovernmental organizations to tackle online violence.

REFERENCES

Adeleke, R. (2020). Digital divide in Nigeria: The role of regional differentials. *African Journal of Science, Technology, Innovation and Development*, https://doi.org/10.1080/20421338.2020.1748335

Akhalbey, F. (2019). Julius Malema Blames Whites for ongoing xenophobia against African migrants in South Africa. *Face2faceafrica*. Retrieved from: face2faceafrica.com/article/julius-malema-blames-whites-for-ongoing-xenophobia-against-african-migrants-in-south-africa-video [Accessed December 20, 2020].

AlDajani, M. I. (2020). *Internet communication technology (ICT) for reconciliation*. Cham: Springer.

Alfredo Acosta, A., & Zia, M. (2020, June 12). Digital transitions in transitional justice. *DeJusticia*. Retrieved from: www.dejusticia.org/en/column/digital-transitions-in-transitional-justice/

Alexandra, S. (2018). Facebook admits it was used to incite violence in Myanmar. *New York Times*. Retrieved from: www.nytimes.com/2018/11/06/technology/myanmar-facebook.html [Accessed October 5, 2019].

Alliance for Peacebuilding. (2012). *Peacebuilding 2.0: Mapping the boundaries of an expanding field*. Washington, DC: United States Institute of Peace. Fall 2012.

Almutawa, A. (2020). Designing the organisational structure of the UN cyber peacekeeping team. *Journal of Conflict & Security Law*, *25*(1), 117–147. https://doi.org/10.1093/jcsl/krz024

Barberá, P. (2020). Social media, echo chambers, and political polarization. In N. Persily, & J. Tucker (Eds.), *Social media and democracy: The state of the field, prospects for reform* (pp. 34–55). Cambridge: Cambridge University Press.

BBC News Mundo. (2019). Álvaro Uribe Denuncia Que Su Cuenta De Twitter Fue "Bloqueada" Durante La Jornada Del Paro Nacional En Colombia. *BBC Mundo*. Retrieved from: www.bbc.com/mundo/noticias-america-latina-50511205 [Accessed September 21, 2020].

Belloni, R., & Moro, F. N. (2019). Stability and Stability Operations: Definitions, Drivers, Approaches. *Ethnopolitics*, *18*(5), 445–461. https://doi.org/10.1080/17449057.2019.1640503

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Berman, E., Felter, J. H., & Shapiro, J. N. (2020). *Small Wars, Big Data: The Information Revolution in Modern Conflict*. Princeton, NJ: Princeton University Press.

Borris, E. R. (2002). Reconciliation in post conflict peacebuilding: Lessons learned from South Africa? *Second track/citizens' diplomacy: concepts and techniques for conflict transformation*. Lanham, MD and Oxford, 161–181.

Bosch, T. (2017). Twitter activism and youth in South Africa: The case of# RhodesMustFall. *Information, Communication & Society*, *20*(2), 221–232.

Bradshaw, S., & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Project on Computational Propaganda. Oxford Internet Institute. University of Oxford. Retrieved from: comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf

Brzoska, M., Ehrhart, H.-G., & Narten, J. (Eds.). (2011). *Multi-stakeholder security partnerships: A critical assessment with case studies from Afghanistan, DR Congo and Kosovo*. Baden-Baden: Nomos.

Calderaro, A., & Craig, A. J. (2020). Transnational governance of cybersecurity: Policy challenges and global inequalities in cyber capacity building. *Third World Quarterly*, *41*(6), 917–938.

Carter, W. (2013). War, peace and stabilisation: Critically reconceptualising stability in Southern Afghanistan. *Stability: International Journal of Security & Development*, *2*(1), 15–35.

Cederman, L. E., & Vogt, M. (2017). Dynamics and logics of civil war. *Journal of Conflict Resolution*, *61*(9), 1992–2016.

Chenou, J. M., Chaparro-Martínez, L. P., & Mora Rubio, A. M. (2019). Broadening conceptualizations of transitional justice through using technology: ICTs in the context of justicia y Paz in Colombia. *International Journal of Transitional Justice*, *13*(1), 92–104.

Chenou, J. M., & Cepeda-Másmela, C. (2019). # NiUnaMenos: Data Activism from the Global South. *Television & New Media*, *20*(4), 396–411.

Chenou, J. M., & Radu, R. (2019). The "right to be forgotten": Negotiating public and private ordering in the European Union. *Business & Society*, *58*(1), 74–102.

Chiumbu, S. (2012). Exploring mobile phone practices in social movements in South Africa– the Western Cape Anti-Eviction Campaign. *African Identities*, *10*(2), 193–206.

Choucri, N., & Clark, D. D. (2018). *International relations in the cyber age: The co-evolution dilemma*. Information Policy.

Cini, L. (2019). Disrupting the Neoliberal university in South Africa: The# FeesMustFall Movement in 2015. *Current Sociology*, *67*(7), 942–959.

Collins, A. (2020). Critical human security and cyberspace: Enablement besides constraint. In M. Salminen, G. Zojer, & K. Hossain (Eds.), *Digitalisation and human security. A multi-disciplinary approach to cybersecurity in the European High North* (pp. 83–109). Cham: Springer.

CONPES. (2016). *Política nacional de seguridad digital*. CONPES No. 3854. Bogotá, DC: Consejo Nacional de Política Económica y Social. Available at: colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3854.pdf

Comninos, A. (2013). *The role of social media and user-generated content in post-conflict peacebuilding*. Washington, DC: World Bank.

DANE. (2020). *Indicadores básicos de TIC en Hogares*. Bogotá, DC: Departamento Administrativo Nacional de Estadísticas. Available at: www.dane.gov.co/index.php/estadisticas-por-tema/tecnologia-e-innovacion/tecnologias-de-la-informacion-y-las-comunicaciones-tic/indicadores-basicos-de-tic-en-hogares

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Dejusticia, Fundación Karisma, & Privacy International. (2017). *The right to privacy in Colombia stakeholder report universal periodic review 30th session – Colombia*. Retrieved from: uprdoc.ohchr.org/uprweb/downloadfile.aspx?filename=5412&file=English Translation

Deibert, R. (2018). Trajectories for future cybersecurity research. In *The Oxford handbook of international security*. Oxford, UK: Oxford University Press.

van Dijk, J. (2020). *The digital divide*. Hoboken, NJ: John Wiley & Sons.

Dunn Cavelty, M. (2013). From cyber-bombs to political fallout: Threat representations with an impact in the cyber-security discourse. *International Studies Review*, *15*(1), 105–122.

du Toit, F. (2017). A broken promise? Evaluating South Africa's reconciliation process twenty years on. *International Political Science Review*, 38(2), 169–184. https://doi.org/10.1177/0192512115594412

Duncombe, C. (2019). The politics of Twitter: Emotions and the power of social media. *International Political Sociology*, *13*(4), 409–429.

Franklin, M. I. (2019). Human rights futures for the internet. In B. Wagner, M. Kettemann, & K. Vieth (Eds.), *Research handbook on human rights and digital technology* (pp. 5–23). Cheltenham, UK: Edward Elgar Publishing.

Geldenhuys, J., & Kelly-Louw, M. (2020). Demistifying hate speech under the PEPUDA. *Potchefstroom Electronic Law Journal*, *23*, 1–50.

Global Partners Digital. (2013). *Internet governance. Towards greater understanding of global south perspectives*. May 2013 Report. London: Global Partners Digital.

Gohdes, A. R. (2018). Studying the internet and violent conflict. *Conflict Management and Peace Science*, *35*(1), 89–106.

Hartzell, C., Hoddie, M., & Rothchild, D. (2001). Stabilizing the peace after civil war: An investigation of some key variables. *International Organization*, *55*(1), 183–208.

Himelfarb, S., & Chabalowski, M. (2008). *Media, conflict prevention and peacebuilding: Mapping the edges*. Washington, DC: United States Institute of Peace.

Hochwald, T. (2013). How do social media affect intra-state conflicts other than war? *Connections*, *12*(3), 9–38.

Hoddie, M., & Hartzell, C. (2005). Signals of reconciliation: Institution-building and the resolution of civil wars. *International Studies Review*, *7*(1), 21–40.

Holmes, C. (2019). *What's behind South Africa's xenophobic violence last week?* The Washington Post. www.washingtonpost.com/politics/2019/09/09/whats-behind-south-africas-xenophobic-violence-last-week/

Howard, R. (2002). *An operational framework for media and peacebuilding*. Vancouver, BC: Institute for Media, Policy and Civil Society.

International Telecommunication Union. (2011). *The quest for cyber peace*. Geneva, January 2011. Retrieved from: handle.itu.int/11.1002/pub/803f9a60-en

International Telecommunication Union. (2020). *Statistics*. Available at: www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Kaplan, A. M., & Haenlein, M. (2012). Social media: Back to the roots and back to the future. *Journal of Systems and Information Technology*, *14*(2), 101–104.

Karlsrud, J. (2014). Peacekeeping 4.0: Harnessing the potential of big data, social media, and cyber technologies In J. F. Kremer, & B. Müller (Eds.), *Cyberspace and international relations. Theory prospects and challenges* (pp. 141–160). Springer.

Kerttunen, M., & Tikk, E. (2020). The Politics of stability: Cement and change in cyber affairs. In E. Tikk, & M. Kerttunen (Eds.), *Routledge handbook of international cybersecurity*. Oxford, UK: Routledge.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Krampe, F. (2016). Empowering peace: Service provision and state legitimacy in Nepal's peace-building process. *Conflict, Security & Development*, 16(1), 53–73. https://doi.org/10.1080/14678802.2016.1136138

Kuehn, A. (2014). Extending cybersecurity, securing private internet infrastructure: The US Einstein program and its implications for internet governance. In R. Radu, J. M. Chenou, & R. Weber (Eds.), *The evolution of global internet governance. Principles and policies in the making* (pp. 157–167). Springer.

Kuerbis, B., & Badiei, F. (2017). Mapping the cybersecurity institutional landscape. *Digital Policy, Regulation and Governance*, 19(6), 466–492. https://doi.org/10.1108/DPRG-05–2017-0024

Laouris, Y. (2004). Information technology in the service of peacebuilding: The case of cyprus. *World Futures*, 60(1–2), 67–79. https://doi.org/10.1080/725289197

Levitt, J. (2019, September 4). #SayNoToXenophobia calls for unity as looting and violence rock SA. TimesLIVE. Retrieved from: www.timeslive.co.za/news/south-africa/2019-09-04-saynotoxenophobia-calls-for-unity-as-looting-and-violence-rock-sa/

Lubin, A. (2020). The rights to privacy and data protection under international humanitarian law and human rights Law. Asaf Lubin, the rights to privacy and data protection under international humanitarian law and human rights law. In R. Kolb, G. Gaggioli, & P. Kilibarda (Eds.), *Research handbook on human rights and humanitarian law: Further reflections and perspectives*. Edward Elgar(forthcoming).

Majcin, J. (2018). Social media challenges to peace-making and what can be done about them. *Groningen Journal of International Law*, 6(2), 242–255.

Margetts, H., John, P., Hale, S., & Yasseri, T. (2015). *Political turbulence: How social media shape collective action*. Princeton, NJ: Princeton University Press.

Maschmeyer, L., Deibert, R. J., & Lindsay, J. R. (2020). A tale of two cybers – how threat reporting by cybersecurity firms systematically underrepresents threats to civil society. *Journal of Information Technology & Politics*. Latest articles, 1–20. https://doi.org/10.1080/19331681.2020.1776658

Mason, T. D., & Mitchell, S. M. (Eds.). (2016). *What do we know about civil wars?* Lanham, MD: Rowman & Littlefield.

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. In Proceedings of the 10th ACM conference on web science (pp. 173–182).

Media Monitoring Africa, South African National Editors' Forum, Interactive Advertising Bureau of South Africa, Association for Progressive Communications. (2019). *Universal access to the internet and free public access in South Africa*. A Seven-Point Implementation Plan. September, 2019. Available at: internetaccess.africa/wp-content/uploads/2019/10/UA-Report.pdf

Meyer, D. (2019). Julius Malema's 'dead white man' tweet is hate speech, says SAHRC. *Sowetan Live*. Retrieved from: www.sowetanlive.co.za/news/south-africa/2019-09-17-julius-malemas-dead-white-man-tweet-is-hate-speech-says-sahrc/?fbclid=IwAR33FjAxA4L5inYf4njbd--mvVfcxy_AZMS7yvRa4lJWZeQxftfRFOWjPUQ

Mielke, K., Mutschler, M., & Meininghaus, E. (2020). For a dynamic approach to stabilization. *International Peacekeeping*, 27(5), 810–835. https://doi.org/10.1080/13533312.2020.1733424

Miklian, J., & Schouten, P. (2019). Broadening 'business', widening 'peace': A new research agenda on business and peace-building. *Conflict, Security & Development*, 19(1), 1–13. https://doi.org/10.1080/14678802.2019.1561612

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Mlonzi, Y. (2017). South Africa and internet governance: Are we just ticking a box? *Global information society watch 2017: National and regional internet governance forum initiatives*. Association for Progressive Communications.

Mueller, M. (2017). Is cybersecurity eating Internet governance? Causes and consequences of alternative framings. *Digital Policy, Regulation and Governance*, 19(6), 415–428. https://doi.org/10.1108/DPRG-05-2017-0025

Mullenbach, M. J. (2005). Deciding to keep peace: An analysis of international influences on the establishment of third-party peacekeeping missions. *International Studies Quarterly*, 49(3), 539–540.

Narten, J. (2011). Multi-stakeholder security partnerships: Characteristics, processes, dilemmas and impacts. In M. Brzoska, H.-G. Ehrhart, & J. Narten (Eds.), *Multi-stakeholder security partnerships* (pp. 15–37). Baden-Baden: Nomos.

Nkanjeni, U. (2019). Twitter rules Malema's 'only trust a dead white man' Mugabe tribute not violent, despite outrage. *Sunday Times*. Retrieved from: www.timeslive.co.za/news/south-africa/2019-09-17-twitter-rules-malemas-only-trust-a-dead-white-man-mugabe-tribute-not-violent-despite-outrage/

Nigam, A., Dambanemuya, H. K., Joshi, M., & Chawla, N. V. (2017). Harvesting social signals to inform peace processes implementation and monitoring. *Big Data*, 5(4), 337–355.

Nilsson, D., & Söderberg Kovacs, M. (2011). Revisiting an elusive concept: A review of the debate on spoilers in peace processes. *International Studies Review*, 13(4), 606–626.

Njeru, S. (2009). Information and communication technology (ICT), gender, and peacebuilding in Africa: A case of missed connections. *Peace and Conflict Review*, 3(2), 32–40.

Odendaal, A. (2010). *An architecture for building peace at the local level: A comparative study of local peace committees*. New York: UNDP.

Onuoha, F. (2013). *Boko Haram: Anatomy of a crisis*, 1–91. Bristol, UK: e-international relations press. Retrieved from: https://reliefweb.int/report/nigeria/boko-haram-anatomy-crisis

Paffenholz, T. (Ed.). (2010). *Civil society & peacebuilding: A critical assessment*. Boulder, CO: Lynne Rienner.

Parlevliet, M. (2017). Human rights and peacebuilding: Complementary and contradictory, complex and contingent. *Journal of Human Rights Practice*, 9(3), 333–357. https://doi.org/10.1093/jhuman/hux032

Pernice, I. (2018). Global cybersecurity governance: A constitutionalist analysis. *Global Constitutionalism*, 7(1), 112–141.

Pettersson, T., & Öberg, M. (2020) Organized violence, 1989–2019. *Journal of Peace Research*, 57(4), 597–613.

Petykó, M. (2018). Troll. In B. Warf (Ed.) *The SAGE encyclopedia of the internet* (pp. 880–882). Thousand Oaks, CA: SAGE Publications.

Preventive Action Working Group. (2015). *Multi-stakeholder processes for conflict prevention and peacebuilding: A manual*. The Hague: Global Partnership for the Prevention of Armed Conflict.

Puig Larrauri, H., & Kahl, A. (2013). Technology for peacebuilding. *Stability: International Journal of Security & Development*, 2(3), 1–15. https://doi.org/10.5334/sta.cv

Puig, H. (2019). Social networks: Fuel to conflict and tool for transformation. *Peace in progress*, 36. Barcelona: ICIP. www.icipperlapau.cat/numero36/articles_centrals/article_central_7

Quintero, R., & Solano, Y. (2020). *Estudiar en línea en Colombia es un privilegio*. El Tiempo, June 30, 2020. Available at: www.eltiempo.com/datos/asi-es-la-conexion-a-internet-en-colombia-510592

Ramsbotham, O., Miall, H., & Woodhouse, T. (2016). *Contemporary conflict resolution*, 4th ed., Cambridge, UK: Polity.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

República de Colombia. (2016). *Acuerdo Final para la terminación del conflicto y la construcción de la Paz Estable y Duradera en Colombia*. November 24, 2016.

Rettberg, A. (2007). The private sector and peace in El Salvador, Guatemala, and Colombia. *Journal of Latin American Studies*, 39(3), 463–494.

Rettberg, A. (2016). Need, creed, and greed: Understanding why business leaders focus on issues of peace. *Business Horizons*, 59(5), 481–492.

Rodríguez-Gómez, D., Foulds, K., & Sayed, Y. (2016). Representations of violence in social science textbooks: Rethinking opportunities for peacebuilding in the Colombian and South African post-conflict scenarios. *Education as Change*, 20(3), 76–97.

Robinson, M., Jones, K., Janicke, H., & Maglaras, L. (2019). Developing cyber peacekeeping: Observation, monitoring and reporting. *Government Information Quarterly*, 36(2), 276–293.

Salem, S. (2014). *The 2011 Egyptian uprising. Framing events through the narratives of protesters. Revolution as a process. The case of the Egyptian uprising*. Bremen, Germany: Wiener Verlag für Sozialforschung, 21–47.

Sarkees, M. R., & Wayman, F. W. (2010). *Resort to war: A data guide to inter-state, extra-state, intra-state, and non-state wars, 1816–2007*. Washington, DC: SAGE Publications.

Sarkin, J. (1998). The development of a human rights culture in South Africa. *Human Rights Quarterly*, 20(3), 628–65.

Schia, N. N. (2018). The cyber frontier and digital pitfalls in the Global South. *Third World Quarterly*, 39(5), 821–837.

Scholte, J. A. (2020). Multistakeholderism: Filling the global governance gap? *Global challenges foundation*, April 2020. Retrieved from: globalchallenges.org/wp-content/uploads/Research-review-global-multistakeholderism-scholte-2020.04.06.pdf

Shackelford, S. J. (2019). Should Cybersecurity be a human right: Exploring the shared responsibility of cyber peace. *Stanford Journal of International Law*, 55(1), 155.

Shackelford, S. J. (2020). *Inside the global drive for cyber peace (April 15, 2020)*. Retrieved from SSRN:ssrn.com/abstract=3577161orhttps://doi.org/10.2139/ssrn.3577161

Shandler, R., Gross, M. L., & Canetti, D. (2019). Can you engage in political activity without internet access? The social effects of internet deprivation. *Political Studies Review*, https://doi.org/10.1177/1478929919877600

Shires, J. (2018). Enacting expertise: Ritual and risk in cybersecurity. *Politics and Governance*, 6(2), 31–40.

South African Human Rights Commission. (2012). *SAHRC statement on latest developments regarding the protection of state information bill*. www.sahrc.org.za/index.php/sahrc-media/news-2/item/151-sahrc-statement-on-latest-developments-regarding-the-protection-of-state-information-bill

South African Human Rights Commission. (2016). *Human rights advocacy and communications report 2015–2016*. www.sahrc.org.za/home/21/files/29567%20A4%20adv%20report%20FINAL%20FOR%20PRINT.pdf

South African Human Rights Commission. (2017). *Submission on the cybercrimes and cybersecurity bill [B6-2017]*. www.sahrc.org.za/home/21/files/SAHRC%20Submission%20on%20Cybercrimes%20and%20Cybersecurity%20Bill-%20Aug%202017.pdf

South African Human Rights Commission. (2019). *Stakeholder dialogue on racism and social media in South Africa*. www.sahrc.org.za/home/21/files/Racism%20and%20Social%20Media%20Report.pdf

Spillane, J. (2015). ict4p: Using information and communication technology for peacebuilding in Rwanda. *Journal of Peacebuilding & Development*, 10(3), 97–103.

State Security Agency. (2015). *The national cybersecurity policy framework*. SA Government Gazette No. 39475. December 4, 2015.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

STATSSA. (2020). *General household survey 2018*. Pretoria: Statistics South Africa. Available at: www.statssa.gov.za/publications/P0318/P03182018.pdf

Stauffacher, D., Weekes, B., Gasser, U., Maclay, C., & Best, M. (Eds.). (2011). *Peacebuilding in the information age. Shifting hype from reality*. Geneva: ICT4Peace Foundation.

Stedman, S. J. (1997). Spoiler Problems in Peace Processes. *International Security*, *22*(2), 5–53.

Tanczer, L. M., Brass, I., & Carr, M. (2018). CSIRT s and global cybersecurity: How technical experts support science diplomacy. *Global Policy*, 9(3), 60–66.

Tellidis, I., & Kappler, S. (2016). Information and communication technologies in peacebuilding: Implications, opportunities and challenges. *Cooperation and Conflict*, *51*(1), 75–93. https://doi.org/10.1177/0010836715603752

Tewathia, N., Kamath, A., & Ilavarasan, P. V. (2020). Social inequalities, fundamental inequities, and recurring of the digital divide: Insights from India. *Technology in Society*, *61*(1), 1–11.

Tully, S. (2014). A human right to access the Internet? Problems and prospects. *Human Rights Law Review*, *14*(2), 175–195.

United Nations. (2005). *Tunis commitment on the information society*. WSIS-05/TUNIS/DOC/7-E. 18 November 2005. Retrieved from: www.itu.int/net/wsis/docs2/tunis/off/7.html

Valeriano, B., & Maness, R. C. (2018). International relations theory and cyber security. *The Oxford Handbook of International Political Theory*, 259.

Verdad Abierta. (2020, September 25). *Quienes somos*. Retrieved from: verdadabierta.com/quienes-somos/

Vyas, K. (2020). *Colombian intelligence unit used U.S. equipment to spy on politicians, journalists*. Retrieved from: www.wsj.com/articles/colombian-intelligence-unit-used-u-s-equipment-to-spy-on-politicians-journalists-11588635893

Wallensteen, P. (2018). *Understanding conflict resolution*. Washington, DC: SAGE Publications.

Walter, B. F. (2017). The new civil wars. *Annual Review of Political Science*, *20*, 469–486.

Weimann, G. (2016). The emerging role of social media in the recruitment of foreign fighters. In A. de Guttry, F. Capone, & C. Paulussen (Eds.), *Foreign fighters under international law and beyond* (pp. 77–95). The Hague: TMC Asser Press.

Wilson, J., & Wilson, H. (2009). Digital divide: Impediment to ICT and peace building in developing countries. *American Communication Journal*, *11*(2), 1–9.

Young, O., & Young, E. (2016). Technology for peacebuilding in divided societies: *ICTs and peacebuilding in Northern Ireland. TRANSCOM (Transformative Connections)*.

Zaum, D. (2012). Beyond the "liberal peace". *Global Governance: A Review of Multilateralism and International Organization*, *18*(1), 121–132.

Zeitzoff, T. (2017). How social media is changing conflict. *Journal of Conflict Resolution*, *61*(9), 1970–1991.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

6

Artificial Intelligence in Cyber Peace

Tabrez Y. Ebrahim

1 INTRODUCTION

This chapter examines artificial intelligence (AI, i.e., or mathematical models for representing computer problems and algorithms for finding solutions to these problems) and its impacts on an arms race (i.e., each nation is focused on self-interest in seeking an incremental gain over another for technological superiority of weapons) (Craig & Valeriano, 2016, p. 142). In the absence of cooperation, all nations are worse off than if they would be if they cooperated in some form. This chapter overviews how AI's unique technological characteristics – including speed, scale, automation, and anonymity – could promote an arms race toward cyber singularity (i.e., a hypothetical point where AI achieves Artificial General Intelligence (AGI), that surpasses human intelligence to become uncontrollable and irreversible) (Newman, 2019, p. 8; Priyadarshini & Cotton, 2020). AI technological advancements have generated a good deal of attention about the AI arms race and its potential for producing revolutionary military applications. While the AI arms race has raised implications for cyber peace, a less studied issue is the potential impact on AGI development in cybersecurity, or cyber singularity. While there is some hype and a development time period toward cyber singularity, the results are generally viewed as negative or, at worst, destabilizing or even catastrophic for cyber peace.

Notwithstanding such limitations, there is still huge potential for the use of technological advancements in AI for civilian, consumer-focused applications, and for the inevitable advancements in nations' military and security technologies. Economic competition for AI has already motivated its development and implementation by the private sector. This has contributed to the imbalance of the economic dominance by industrialized countries. Innovative companies and countries that focus on AI development may begin to monopolize AI knowledge and take the lead toward cyber singularity, which could thwart cyber peace. AI has also become an essential component of cybersecurity, as it has become a tool used by both attackers and defenders alike (Roff, 2017). In the future, the more advanced form of AGI, or super technological intelligence, could develop its own understanding of the

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

world and react to it in a rapid and uncontrollable way without human involvement. Advancement toward cyber singularity could present new military capabilities, such as manipulation of data and overcoming other nations' defenses, and transform interactions in cyber conflict. While is difficult to detect or measure the origination or proliferation of AI in cybersecurity, whatever possible cooperation among nations that can be promoted is certainly worth exploring. Thus, this chapter explores how shared governance through talent mobilization in the form of a global AI service corps can offset the negative impact of nation-states' economic competition to develop AGI.

2 BACKGROUND AND CHARACTERIZATION OF AI

The definition of AI varies in context and is a moving target as technology continues to advance (Lemley & Case, 2020, p. 1). The term AI is meant to refer to computer programs that perform mathematically oriented tasks that were generally assumed to require human intelligence (Lefkowitz, 2019). AI can take a variety of forms including logical inference (a form of deduction) and statistical inference (of form of induction or prediction) (Eldred, 2019). Such mathematical techniques are becoming more powerful because of the availability and use of large datasets, easy access to powerful and inexpensive computing resources, and the ability to run new algorithms and solve complex problems using massive parallel computing resources (Firth-Butterfield & Chae, 2018, p. 5; Daly, 2019). Another way to look at the current state of AI is that it has become cheaper and easier to utilize its techniques with more speed, scale, and automation than ever before. Moreover, the platforms of collecting, using, and solving relationships in data can be done anonymously, which presents opportunities for exploitation of consumers in business and nations in cyber conflict.

Technological advancements have always played a crucial role in the context of conflict and peace (Roff, 2016, p. 15). The introduction of information technology presented opportunities to create, move, and process data in ways never seen before, leaving nations with the power to control, defend, secure, and weaponize data. AI performs these tasks better, faster, and with more anonymity than humans, and outperforms ordinary computers and networked systems.

The information technology sophistication of AI allows for disguised and stealth measures, provides for more effective and contextualized threats, and has the potential for amplified human cognitive capabilities in the form of cyber singularity over time (Priyadarshini & Cotton, 2020). Many characteristics of information technology – including its ability to involve multiple actors, attribute challenges, and proliferate across borders – present unprecedented challenges for AI in cyber peace (Geers, 2011, p. 94). Information technology warfare and protection measures in the modern day present unique considerations for AI compared to prior means and methods. In this vein, AI-based information technologies related to cyber peace fall

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

into three primary classifications: (1) information attacks; (2) information anonymity; and (3) information attribution (Reuter, 2020, p. 16, 24–5, 113–14, 117, 279–81). A new classification of manipulation or change by AI, which is increasingly becoming ubiquitous, presents new opportunities for the integration of multiple stakeholder input.

With AI, nations can analyze patterns and learn from them to conduct cyberattacks (i.e., offensive capabilities of AI) and also use these patterns prevent cyberattacks (i.e., defensive capabilities of AI) in more advanced mechanisms than current capabilities. The state of the art AI already allows for discovering of hidden patterns in data and automating and scaling mathematical techniques with data to make predictions (Coglianese & Lehr, 2018, pp. 14–15).

The path toward AGI is especially attractive insofar as it will not seem to require human intervention and will control the information infrastructure in cyber conflicts (Burton & Soare, 2019, pp. 5–6). As the tools, techniques, and software become increasingly intelligent, AI will have greater role in cyber conflict and cyber peace. To assess this path toward AGI and its implications for shared governance, an overview of information security technology and AI's role in information security is necessary as a preliminary matter.

2.1 *Information Security Overview*

The stakes in our national information security debate are high. Information security refers to the hybrid scientific and legal inquiry into defending against all possible third-party attackers and the legal consequences that arise when they cannot. The purpose of information security is to develop and provide technological solutions to prevent the potential for cyberattacks and to minimize the interstate insecurity caused by information technologies (Libicki, 2009, pp. 12–13). Information security technologies have a crucial impact on AI's role in cyber peace, and therefore, it is necessary to have a proper understanding of what these concepts mean and how they may accelerate or decelerate concerns for a path toward a sustainable and secure cyber peace.

Information security is a capricious concept with varying definitions in the legal and policy realms, but it has a more concrete meaning in computer science and technological realms (Reuter, 2020, pp. 17–18). In a technological sense, the cyber world of computerized networks where information technologies are relevant have three layers: (1) a physical layer of infrastructure (including integrated circuits, processors, storage devices, and optical fibers); (2) a software logic layer (including computer programs and stored information that is subject to processing); and (3) a data layer, for which a machine contains and creates information (Tabansky, 2011, p. 77). In order to analyze the relevance of information technology, particularly AI, and its role in cyber peace, it is necessary to understand how these concepts relate to technology characteristics. While conflicts among nations can be carried out in different domains, such as land, sea, air, and space, conflict with the use of

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

information technology infrastructure has the following peculiar characteristics for security implications: (1) many actors can be involved; (2) the identity of the security threat may be unknown due to the challenge of attribution; (3) international proliferation; and (4) its dual-use nature that can be exploited in a variety of ways (Reuter, 2020, pp. 12–13). These characteristics are accounted for in the various defensive and offensive uses of information technology, as subsequently shown.

2.2 *Defensive Information Security Measures*

Defensive protection measures allow for proactive ways to detect and obtain information regarding cyberattacks or intrusion (Chesney, 2020, p. 3). Defending against cyberattackers entails the use of software tools that obfuscate or obscure cyberattackers' efforts (Andress & Winterfeld, 2011, p. 113). A major goal of defensive cyber protection is to prevent critical infrastructure damage which would generate large spillover effects in the wider economy. The defensive cyber protection approach seeks to: (i) minimize unauthorized access, disruption, manipulation, and damage to computers and (ii) mitigate the harm when such malicious activity occurs to computers. In so doing, information security seeks to preserve the confidentiality, integrity, and availability of information (Tabansky, 2011, p. 81).

Approaches fall into two general categories: proactive measures (also known as preventative techniques, which can block efforts to reach a vulnerable system via firewalls, access controls, and cryptographic protection) and deterrence measures (that increases the effort needed by an adversary, and includes many types of security controls) (Ledner et al., 2009, pp. 6–7, 9–10). In either approach, the goal is to prevent unauthorized access to a computer system by the use of technological methods to identify an unauthorized intrusion, locate the source of the problem, assess the damage, prevent the spread of the damage, and reconstruct damaged data and computers (Reuter, 2020, pp. 22, 280–283). Deterrence, mitigation, and preventative strikes with the use of information technology include application security, attack detection and prevention, authorization and access control, authentication and identification, logging, data backup, network security, and secure mobile gateways.

2.3 *Offensive Information Security Measures*

While defensive measures and technology can deter and mitigate the consequences of unauthorized access of computers and networks, limiting unauthorized access may not achieve cyber policy goals. Offensive measures, which are considered lawful but unauthorized, refer to penetrating or interfering with another system and can include mechanisms that allow for impersonation of trusted users and faster attacks with more effective consequences (Dixon & Eagan, 2019). Such offensive measures are one of many ways that nations can utilize cyber power to destroy or disable an adversary's infrastructure (Voo et al., 2020). Nations seek to achieve cybersecurity

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

in order to bend the other side's will or to manage the limiting the scope of the other side's efforts, and can do so, via deliberate provocation or through escalation via offensive measures or cyberattacks (Jensen, 2009, pp. 1536–1538). A common mechanism for cyberattacks is a computer network attack, wherein actions are taken through the use of information technology and computer networks to disrupt, deny, degrade, or destroy information on computers and networks, and can electronically render useless systems and infrastructures (Andress & Winterfeld, 2011, pp. 110–113).

The increasing power of computers, proliferation of data, and advancements in software for AI capabilities presents many new applications of offensive measures. To demonstrate that AI is a rapidly growing field with potentially significant implications for cyber peace, several technological examples are provided to show the direct or indirect impact of such technological advancement on the need for *shared governance of a global service AI corps.*

Attack means and methods include malware, ransomware, social engineering, advanced persistent threats, spam, botnets, distributed denial of service, drive-by-exploits and exploit kits, identity theft, and side channel attacks. Such cyberattacks include the intrusion of the digital device with some sort of malware that initiates the communication between the attacking computing and the intruded device. The reasons for initiating such offensive measures include preventing authorized users from accessing a computer or information service (termed a denial-of-service attack) destroying computer-controlled machinery, or destroying or altering critical data and, in doing so, can affect artifacts connected to systems and networks (such as cyber-physical devices, including generators, radar systems, and physical control devices for airplanes, cars, and chemical manufacturing plants). Cyberattack mechanisms include the use of malware installation (sometimes combined with disruptive code and logic bombs), creation of botnets (that refer to a group of infected and controlled machines that send automated and senseless reports to a target computer), and installation of ransomware (that encrypts a device) (Reuter, 2020, pp. 16, 24–5, 113–14, 117, 140, 279–81). Malware refers to malicious software, which can attack, intrude, spy on, or manipulate computers. Botnets are made up of vast numbers of compromised computers that have been infected with malicious code and can be remotely controlled through Internet-based commands. Ransomware refers to malicious software that is installed on a computer, network, or service for extortion purposes, by encrypting the victim's data or systems and making them unreadable such that the victim has to submit a monetary payment for decrypting files or regaining access.

2.4 *Information Security Linkage to Artificial Intelligence*

Technological development, particularly in the rapidly developing information technology realm, plays a crucial role in questions regarding cyber peace. Information technology is becoming omnipresent in the cases of resilience and of

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

managing cyber conflicts. As the interdisciplinary field of cyber peace links more with technology, it is crucial to consider the ways that information technology assists and supports peace processes, as well as be cognizant of ways it can be a detriment.

Ever since information technology has created, moved, and processed data, the security of the data encountered challenges with policy and conflict resolution. In recent years, as advancements in information technology have increased connectivity, collaboration, and intelligence, these issues have become even more important. Information technology concerns information sharing and deterrence and implicates security concerns. As such, information technology security involves the preservation of confidentiality, integrity, availability, authenticity, accountability, and reliability. Relatedly, information technology can manipulate and anonymize data, and this feature can be used for a cyberattack (Gisel & Olejnik, 2008, pp. 14–17). The implication of this capability is attribution challenges. Attribution refers to the allocation of a cyberattack to a certain attacker toward providing real-world evidence for unveiling the identity of the attacker. AI makes it easier to identify or attribute a cyberattacker since it analyzes significantly higher number of attack indicators and discovers patterns (Payne, 2018).

AI is poised to revolutionize cyber technological use in cyber peace, by providing faster, more precise, and more disruptive and anomalous capabilities (Stevens, 2020, pp. 1, 3, 4). AI can analyze data and trends to identify potential cyberattacks and provide offensive countermeasures to such attacks (Padrón & Ojeda-Castro, 2017, p. 4208). Moreover, AI presents the most powerful defensive capability in cybersecurity (Haney, 2020, p. 3). While AI presents new technological capabilities to cyber conflict, it raises new considerations of what it might mean for human control, or lack thereof, and how it may help or hinder risks (Burton & Soare, 2019, pp. 3–4). AI capabilities can undermine data integrity and present stealthy attacks that cause trust in organizations to falter and lead to systemic failures (Congressional Research Service, 2020, Summary). Nations could use AI to penetrate another nation's computers or networks for the purposes of causing damage or disruption through manipulation and change (Taddeo & Floridi, 2018, pp. 1–2).

From an offensive standpoint, AI presents new considerations for cyber conflict, such as new manipulation or change capabilities that can allow for expert compromise of computer systems with minimal detection (Burton & Soare, 2019, pp. 9–10). Adversarial AI impacts cyber conflict in three ways, including impersonation of trusted users, blending in the background by disguise and spreading itself in the digital environment, and faster attacks with more effective consequences. These capabilities provide motivation for the "defend forward" strategy of a preemptive instead of a reactive response to cyberattacks (Kosseff, 2019, p. 3).

Additionally, AI makes deterrence possible since its algorithms can identify and neutralize the source without necessarily identifying the actor behind it, which makes it easier to thwart attacks. AI capabilities allow for going to the forefront of

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

the cause or the conflict to analyze data and trends to identify potential attacks and provide countermeasures to such attacks.

3 PATH TOWARD AGI AND IMPLICATIONS FOR CYBER SINGULARITY

The technological development and advancement of AI presents challenges and lessons for governance frameworks. Social science research has been applied toward addressing governance gaps with AI, including polycentric governance and the resulting implications for policymakers (Shackelford & Dockery, 2019, pp. 6–7; Shackelford, 2014, pp. 2, 4–5).

There is no single definition of AGI, but the general consensus is that AGI refers to machines gaining intelligence that is greater than that of humans (Payne, 2018). When AGI is applied to cybersecurity, it has been termed cyber singularity, which presents superintelligence and amplification of human cognitive capabilities in cyberspace. The path toward AGI involves advancements in the form of a technological tool in a classical scenario and in the application of such a tool in novel situations.

The race to AGI involves the development of tools (mathematical techniques and software) used in classical cyber offense and cyber defense scenarios, but with increasing intelligence (Burton & Soare, 2019, pp. 5–6). These represent technological attacks on computer networks, data, and infrastructure. While achieving AGI is a futuristic concept, advancements in sensory perception and natural language understanding will help transform AI into AGI and present new offensive and defensive capabilities in cyber peace. The offensive capabilities of AGI could involve sabotaging data, masking and hiding it being a cyberattack, and engaging in changing behaviors and contextualizing its threats. The defensive capabilities of AGI could involve automatically scanning for vulnerabilities in computer networks, gathering intelligence through the scanning of computer systems, and improving existing software and scripts. In both the offensive and defensive realm, AGI could manipulate humans or detect when humans were being manipulated and respond accordingly. Similar to an advanced form of psychological manipulation of behavioral advertising, AGI could conduct sophisticated manipulation of human decision-making in the midst of a cyber conflict and, in doing so, could amplify points of attack, coordinate resources, or stage attacks at scale (National Science & Technology Council, 2020, p. 7).

The race toward AGI also involves application of such tools in novel forms pertaining to cybersecurity (Geist, 2016; Cave & ÓhÉigeartaigh, 2018). In additional to technological attacks on computer networks, data, and infrastructure, AGI could be applied to psychological manipulation in society to shape information in the political realm, the Internet, and social media with national cybersecurity implications. In the context of cybersecurity, AGI, as applied to manipulation of people with societal impact, includes shaping public understanding and political action that

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

impacts national cybersecurity policy. Unlike the scenario of AGI as a technological tool, in a related manner, AGI as socio-political manipulator can provide an automated mass deception or mass data collection that implicates national cybersecurity and global perspectives. While not as direct an impact as a technological attack on computer networks, data, and infrastructure, this form of AGI provides manipulative messaging and interference in media, politics, and the public sphere, akin to the profiling and data analysis methods implemented in the Cambridge Analytica scandal.

In addition to the advancement of AI toward AGI for use as a technological tool, and its application to shape the socio-political information realm, AGI technological advancement in the form of cyber singularity would necessitate transformation of warfare approaches (Ivey, 2020, p. 110; O'Hanlon, 2018). Cyber singularity, or the hypothetical point of AGI, becomes uncontrollable and irreversible in the cybersecurity realm and implicates international initiatives and policies (Priyadarshini & Cotton, 2020). The literal interpretation of cyber singularity concerns targeting weapons advancement with an offset strategy, or achieving technological superiority for deterrence effects. Similar to past offset strategies with nuclear weapons and information surveillance and stealth weapons, AGI for cyber singularity represents the next offset strategy. The strategic development and use of modern algorithms, data, and information on computer networks in the path toward AGI is critical in the AI arms race. In this sense, the world is at a critical stage in the strategic use of data and control of information on computer networks. As nations seek AGI capabilities in the AI arms race, policies that promote its development are of critical importance. A shared governance approach in some form should consider ways to offset the negative impact of nation-states' economic competition to develop AGI.

4 SHARED GOVERNANCE OF A GLOBAL SERVICE AI CORPS

The idea about the path toward AGI and implications of cyber singularity is that it might be possible to create a computational machine that vastly outperforms humans in cognitive areas of cybersecurity. Whereas current state of the art AI can apply to limited cybersecurity domains, AGI could also learn and expand into more cyber domains. The potential for AGI is speculative and the idea of cyber singularity is fuzzy since it is unclear what technologies are necessary for its realization. Thus, with an unclear understanding of the likelihood and function of cyber singularity, the technological development pathway raises a host of questions. By contrast, nations could foreseeably control governance strategies in relation to AGI. One potential option – that this chapter prescribes – is directing talent and human resources toward cooperation.

Nations that direct human capital resources in this way would allow for exerting control of human behavior in the arms race toward AGI and implications toward cyber singularity. Currently, there is a "brain drain" of AI talent that is largely employed

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

by the private sector (Andress & Winterfeld, 2011, p. 248; Congressional Research Service, 2009, p. 22). A commission that recruits, develops, and retains AI talent, such as in the form of a reserve corps, could help to equalize the playing field in the AI arms race and transform governance away from state-centric approaches to AI. The facilitation of early global coordination among multiple stakeholders with common interests and sharing of best practices could prevent global catastrophic cybersecurity risks (Newman, 2019, p. 4). Such a multistakeholder policy toward AI development represents a system flexible enough to adapt to new challenges and realities in a global system and toward cyber peace, potentially even forming the backbone of a Cyber Peace Corps (Shackelford, 2017). Given that AI technological development toward AGI has been under the purview of nations, the solution to the problem of an AI arms race toward cyber singularity needs to be rooted through multilateral networks.

The AI arms race has largely been framed by its economic impact rather than in shared governance structures. As a result, industrialized countries with strong software industries have continued to develop AI tools that have skewed the AI arms race. As AI and data implicate economic wealth and political influence, cyber peace policy conversations will need to consider the role and advancement of AI. The greatest threat to and the greatest opportunity for cyber peace could be AI technology, rather than other forces in the nations themselves.

REFERENCES

Andress, J., & Winterfeld, S. (2011). *Cyber warfare*. Elsevier.

Burton, J., & Soare, S. (2019). *Understanding the strategic implications of the weaponization of artificial intelligence* [Manuscript]. 11th international conference on cyber conflict. Tallinn, Estonia. https://ccdcoe.org/uploads/2019/06/Art_14_Understanding-the-Strategic-Implications.pdf

Cave, S., & ÓhÉigeartaigh, S. (2018, December). *An AI race for strategic advantage: Rhetoric and risks*. AIES '18: Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society. New Orleans, LA, USA. https://doi.org/10.1145/3278721.3278780

Chesney, B. (2020, March 2). *Cybersecurity law, policy, and institutions, v.3.0.*

Congressional Research Service. (2020, August 26). *Artificial intelligence and national security.*

Congressional Research Service. (2009, March 17). *Information operations, cyberwarfare, and cybersecurity: Capabilities and related policy issues.*

Coglianese, C., & Lehr, D. (2018, November 9). Transparency and Algorithmic Governance. *Administrative Law Review*, 71(1), 1–56.

Craig, A. & Valeriano, B. (2016). *Conceptualising cyber arms races* [Manuscript]. 8th International conference on cyber conflict. Tallinn, Estonia. https://ccdcoe.org/uploads/2018/10/Art-10-Conceptualising-Cyber-Arms-Races.pdf

Daly, A. (2019, June 5). *Artificial intelligence governance and ethics: Global perspectives.* https://arxiv.org/ftp/arxiv/papers/1907/1907.03848.pdf

Dixon, W., & Eagan, N. (2019, June 19). *3 Ways AI will change the nature of cyber attacks*. World Economic Forum. www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Eldred, C. (2019, October). *AI and domain knowledge: Implications of the limits of statistical inference*. Berkeley Roundtable on International Economics. https://brie.berkeley.edu/sites/default/files/ai_essay_final_10.15.19.pdf

Firth-Butterfield, K., & Chae, Y. (2018, April). *Artificial intelligence collides with patent law*. World Economic Forum. www3.weforum.org/docs/WEF_48540_WP_End_of_Innovation_Protecting_Patent_Law.pdf

Geers, K. (2011, January 1). *Strategic cyber security*. NATO Cooperative Cyber Defence Centre for Excellence.

Geist, E. M. (2016, August 15). It's already too late to stop the AI arms race—we must manage it instead. *Bulletin of the Atomic Scientists*, 72(5), 318–321. https://doi.org/10.1080/00963402.2016.1216672

Gisel, L., & Olejnik, L. (2008, November 14–16). *The potential human cost of cyber operations* [Manuscript]. ICRC Expert Meeting. Geneva, Switzerland. www.icrc.org/en/document/potential-human-cost-cyber-operations

Haney, B. S. (2020). Applied artificial intelligence in modern warfare & national security policy. *Hastings Science and Technology Journal*, *11*(1), 61–100.

Ivey, M. (2020). The ethical midfield in artificial intelligence: Practical reflections for national security lawyers. *The Georgetown Journal of Legal Ethics*, 33(109), 109–138. www.law.georgetown.edu/legal-ethics-journal/wp-content/uploads/sites/24/2020/01/GT-GJLE190067.pdf

Jensen, E. T. (2009). Cyber warfare and precautions against the effects of attacks. *Texas Law Review*, 88(1533), 1534–1569.

Kosseff, J. (2019). *The countours of 'Defend Forward' under international law*, 2019 11th International Conference on Cyber Conflict (CyCon) 900, 1–13.

Ledner, F., Werner, T., & Martini P. (2009). Proactive botnet countermeasures – An offensive approach. In C.Czosseck & K.Geers (Eds.), *The virtual battlefield*: *Perspectives on cyber warfare* (pp. 211–225). 10.3233/978-1-60750-060-5-211

Lefkowitz, M. (2019, September 25). *Professor's perceptron paved the way for AI – 60 years too soon*. Cornell Chronicle. https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon

Lemley, M. A., & Case, B. (2020). You might be a robot. *Cornell Law Review*, *105*(287), 287–362.

Libicki, M. C. (2009). *Cyberdeterrence and cyberwar*. RAND Corporation.

National Science & Technology Council. (2020, March). Networking & Information Technology Research and Development Subcomittee and the Machine Learning & Artificial Intelligence Subcommittee. Artificial Intelligence and Cybersecurity: Opportunities and Challenges, Technical Workshop Summary Report.

Newman, J. C. (2019, February). *Towards AI security: Global aspirations for a more resilient future*. Center for Long-Term Cybersecurity.

O'Hanlon, M. E. (2018, November 29). *The role of AI in future warfare*. Brookings. www.brookings.edu/research/ai-and-future-warfare/

Padrón, J. M., & Ojeda-Castro, Á. (2017, June). Cyberwarfare: Artificial intelligence in the frontlines of combat. *International Journal of Information Research and Review*, 4(6), 4208–4212.

Payne, K. (2018). *Artificial intelligence: A revolution in strategic affairs?* International Institute for Strategic Studies.

Priyadarshini, I., & Cotton, C. (2020, May 6). Intelligence in cyberspace: The road to cyber singularity. *Journal of Experimental & Theoretical Artificial Intelligence*. https://doi.org/10.1080/0952813X.2020.1784296

Reuter, C. (2020). *Information technology for peace and security*. Springer.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Roff, H. M. (2016, March). *Cyber peace, new America*. Cybersecurity Initiative.

Roff, H. M. (2017, August 1–3). *Cybersecurity, artificial intelligence, and nuclear modernization* [Workshop]. Cyberwarfare and Artificial Intelligence. University of Iceland, Reykjavik, Iceland.

Shackelford, S. J. (2014, April 16). Governing the final frontier: A polycentric approach to managing space weaponization and debris. *American Business Law Journal*, 51(2), 429–513.

Shackelford, S. J. & Dockery, R. (2019, October 30). Governing AI. *Cornell Journal of Law and Policy*. Advanced online publication.

Stevens, T. (2020, March 31). Knowledge in the grey zone: AI and cybersecurity. *Journal of Digital War*. https://doi.org/10.1057/s42984-020-00007w

Tabansky, L. (2011, May). Basic concepts in cyber warfare. *Military and Strategic Affairs*, 3(1), 75–92.

Taddeo, M., & Floridi, L. (2018, April 16). Regulate artificial intelligence to avert cyber arms race. *Nature*, 556(7701), 296–298. https://doi.org/10.1038/d41586-018-04602-6

Voo, J., Hemani, I., Jones, S., DeSombre, W., Cassidy, D., & Schwarzenbach, A. (2020, September). *National Cyber Power Index 2020*. Belfer Center for Science and International Affairs, Cambridge, Tech. Rep. September 2020 [Online]. Available: www.belfercenter.org/sites/default/files/2020-09/NCPI_2020.pdf

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

PART III

Lessons Learned and Looking Ahead

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

7

Contributing to Cyber Peace by Maximizing the Potential for Deterrence

Criminalization of Cyberattacks under the International Criminal Court's Rome Statute

Jennifer Trahan[*]

1 INTRODUCTION

This chapter examines how a cyberattack (i.e., a cyber-enabled malicious activity) that has consequences similar to a kinetic or physical attack (causing serious loss of life or physical damage) could be encompassed within the crimes that may be prosecuted before the International Criminal Court (ICC). The chapter overviews when and how such a cyberattack could fall within the ambit of the ICC's crimes – genocide, crimes against humanity, war crimes, and the crime of aggression. The chapter additionally acknowledges some of the limitations as to which attacks would be encompassed, given, particularly, the gravity threshold of the ICC's Rome Statute, as well as the hurdle of proving attribution by admissible evidence that meets the standard of proof beyond a reasonable doubt. Notwithstanding such limitations, there is still potential for use of the Rome Statute to encompass a limited subset of cyberattacks. Increased awareness of this previous largely overlooked potential could possibly contribute to deterring such crimes, as could prosecution of those cases of cyberattacks that meet the standard of proof by required by the ICC Rome Statute. While it is very difficult to measure the deterrent impact of tribunals and international criminal law, whatever possible deterrence that can be created is certainly

[*] Clinical Professor, NYU Center for Global Affairs and Director of the Concentration in International Law & Human Rights. A more extensive version of the topics addressed herein will appear in Jennifer Trahan, "Criminalization of Cyber-Attacks under the International Criminal Court's Rome Statute" (Trahan, forthcoming). The author thanks Pano Yannakogeorgos for aiding her understanding of cyber operations. She also greatly benefitted from discussions at meetings of the Council of Advisers on the Application of the Rome Statute to Cyberwarfare. The author additionally benefitted from workshopping her chapter at the April 17 and September 25, 2020 workshops hosted by the Ostrom Workshop Program on Cybersecurity and Internet Governance at Indiana University, and particularly appreciates the comments of her discussant, Asaf Lubin. She also benefitted from workshopping the chapter at the June 29–30, 2020 International Criminal Court Scholars' Forum, and particularly appreciates the comments of her discussant, Elies van Sliedregt, and those of Matthew E. Cross, Erin Lovall, Kara McDonald, and Samantha Wynne who provided research assistance.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

worth maximizing. This chapter explores how international criminal law could potentially contribute to the goal of reaching a state of "cyber peace." Admittedly, the Rome Statute would not encompass the vast number of cybercrimes that occur, as it would only cover the more severe cyberattacks, such as those inflicting serious loss of life or significant physical damage; however, the Rome Statute *does* have applicability in this area to cover at least a limited subset of cyber operations, and this potential should be explored and utilized. The ICC can only help contribute to deterrence and cyber peace if the ability of the ICC to prosecute certain cyberattacks becomes acknowledged and well known.

2 BACKGROUND

Cyberattacks can take a variety of forms including those aimed at data theft (stealing corporate information) (Griffiths, 2015; as cited in Jensen, 2017, p. 736, n. 6), extortion, the spreading of false information (Greenfield, 2013; as cited in Jensen, 2017, p. 736, n. 7), manipulation of elections (Hathaway et al., 2012, p. 819; Ohlin, 2020), breach of government computers in an effort to steal state secrets (O'Hare, 2016; as cited in Jensen, 2017, p. 737, n. 8,), as well as denial of service attacks (U.S. Department of Homeland Security and the Cybersecurity and Infrastructure Security Agency [CISA], 2019). Such attacks sometimes proliferate beyond their intended targets to impact information technology (IT) systems around the globe, as with the infamous NotPetya attack (Greenberg, 2018; Efrony & Shany, 2018, pp. 628–631). Further, the software code used in cyberattacks may also be "repurposed" by others (Bellovin et al., 2017). Fortunately, to date, cyberattacks and cybercrimes have not resulted in widespread devastation causing loss of life and, to the extent physical damage has resulted, such damage has occurred more to data, computer hardware and software and, in one instance, to centrifuges at a nuclear facility (the Stuxnet attack). Cyberattacks have also caused massive economic losses (Greenberg, 2018) and compromised the personal information of millions of individuals (Jensen, 2017, p. 737).

The use of cyber technology to date, however, makes clear that a much more catastrophic cyberattack could occur. States (on their own, or in conjunction with nonstate actor/hacker groups) now have the capacity to combine cyber weapons with conventional weapons into a "blended attack," such as occurred in Ukraine (Greenberg, 2018) and Georgia (ICC Forum, 2018). A number of war crimes that could be committed during a conventional armed conflict could now, potentially, also be committed through the use of cyberattacks or through both cyber and conventional means. Absent a state of armed conflict, cyberattacks meeting the requirements of, *inter alia*, a widespread or systematic attack against a civilian population could fall within the ambit of "crimes against humanity." For example, the technological capacity to disable air traffic controls exists, causing a "cyber 9/11" – perpetrated, for example, by nonstate actors (cyber criminals or bands of hackers). A cyberattack could similarly target computer systems that control train traffic, nuclear

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

facilities (Greenberg, 2017a), hospitals (Glaser, 2017; Mačák et al., 2020), power grids (Greenberg, 2017b, 2018), and other critical infrastructure – or, for example, a dam located upstream from a major city (Berger, 2016). It is this narrow subset of cyberattacks – causing serious loss of life and/or physical destruction – *not* the vast number of cybercrimes being perpetrated – that this chapter addresses.

While one hopes that a large-scale cyberattack, or even a more limited one that causes serious loss of life or damage to physical objects, will never reach fruition, it is simultaneously helpful to acknowledge that should such an attack occur, it potentially could be covered by one or more of the crimes provided for under the ICC's Rome Statute. States, particularly Rome Statute States Parties, could also incorporate Rome Statute crimes into their domestic criminal codes and statutory law (if they have not already done so), and/or develop additional laws criminalizing cyberattacks and/or cybercrimes. Should they do so, domestic definitions of the crimes could be more broadly formulated than their Rome Statute counterparts and have greater jurisdictional reach; thus, the limitations of the Rome Statute discussed in this chapter do not necessarily carry over to domestic jurisdictions. While the international community could also create a new international criminal tribunal to deal exclusively with cybercrime and cyberattacks, in light of the apparently unwillingness to create new criminal tribunals[1] this chapter focuses on the permanent international criminal tribunal that already exists, the ICC.

3 MAXIMIZING THE POTENTIAL FOR CYBER PEACE THROUGH DETERRENCE

The goal of the present chapter is not only to make the case for ICC cyber prosecutions should a horrific attack occur but to increase awareness of the potential for ICC prosecutions in order to maximize the potential for deterrence as a means to contribute to achieving a state of "cyber peace" (see Shackelford, 2017, p. 8, defining "cyber peace"). It is important that the cyber domain is not seen as unfettered by the rule of law, when it is in fact subject to numerous bodies of international law (UK Government, 2018; Koh, 2012, p. 3), including international humanitarian law (aka the laws of war) (Schmitt, 2017, Rule 80; "The Paris Call," 2018), international human rights law (Schmitt, 2017, Rule 35), as well as the use of force norms contained within the UN Charter as supplemented by customary international law (Schmitt, 2019 (citing position of France, and the UK Government, 2018, noting also that Russia and China accept that the UN Charter applies in cyberspace)). The more well acknowledged it is that international humanitarian law and international human rights law apply in the cyber domain, the easier it is to make the case

[1] Recently, for example, the international community has created three investigative mechanisms – to investigate crimes committed in Syria, Myanmar, and Iraq (if perpetrated by the so called "Islamic State" (ISIL) – but has not created tribunals for the prosecution of those crimes (see Trahan, 2021).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

that certain cyberattacks are covered under international criminal law. Even if the application of current bodies of international law to cyberattacks may not prove an "elegant fit," it is imperative to utilize the laws that exist and/or develop additional laws (cf. Rona, 2003, p. 60, arguing International Humanitarian Law (IHL) should apply to the "war on terror" even if "not an elegant fit").

Significant academic literature exists on the subject of whether international criminal law can play a deterrent role and whether the existence of the various *ad hoc* and hybrid criminal tribunals has contributed to deterrence and/or the ICC can do so.[2] Various scholars take a pessimistic stance as to the potential of tribunals to deter atrocity crimes (McAllister, 2019–20, p. 85, n. 2, categorizing scholars as "deterrence pessimists"). Yet, increasingly, there are scholars whose studies yield positive results (McAllister, 2019–20, p. 85, n. 4, categorizing scholars as "deterrence optimists"). For example, a recent study, based on over 200 interviews, demonstrates that Macedonian Armed Forces, during the 2001 conflict in Macedonia, considered the existence of the International Criminal Tribunal for the former Yugoslavia (ICTY) when deciding their actions (specifically, whether any could be viewed as war crimes), and this deterred violence against civilians (McAllister, 2019–20; see also Schense & Carter, 2017). Similar studies show some deterrence created by the existence of the ICC (Jo & Simmons, 2016; Hillebrecht, 2016; Human Rights Watch, 2009, Ch. IX).

It is worth noting that domestic criminal law also does not fully deter domestic crimes; yet states nonetheless criminalize crimes, from murder to insider trading. So too with international criminal law. As Brierly observes: "States often violate international law, just as individuals often violate municipal law" (Brierly, 1944, pp. 4–5). Clearly, the field of international justice has not yet fully deterred crimes such as genocide, crimes against humanity, or war crimes, as these crimes still occur far too often. Furthermore, it is also notoriously hard to prove a negative – that is, that crimes have *not* occurred due to the deterrent impact of tribunals or international criminal law – so there could actually be more deterrence than can be conclusively demonstrated. Yet, the case that one should *not* criminalize atrocity crimes is generally not made; clearly, whatever role deterrence can play is worth maximizing, and if international criminal laws and tribunals are incapable of deterring or not fully capable of doing so, then at least the laws exist whereby the crimes may be prosecuted. In short, international criminal law is one of the tools at the disposal of those working in the field of international justice, and while it may not fully deter, any deterrence potential is useful. As Guido Acquaviva writes: "international criminal institutions" that "strengthen[] the rule of law and pursu[e] individual criminal

[2] The *ad hoc* tribunals refer to the International Criminal Tribunal for the former Yugoslavia (ICTY) and the International Criminal Tribunal for Rwanda (ICTR). The "hybrid tribunals" include the Special Court for Sierra Leone, the Extraordinary Chambers in the Courts of Cambodia, the hybrid War Crimes Chamber of the Court of Bosnia and Herzegovina in Sarajevo (State Court), the Special Tribunal for Lebanon, and the Kosovo Specialist Chambers.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

responsibility" "can increase awareness of the primary rules … among the general public and … foster compliance with the law and therefore, indirectly, general deterrence" (Acquaviva, 2014, p. 786).

The United States, for instance, suggests that the prosecution of cyberattacks *can* change behavior. Kristen Eichensehr explains:

> One of the most often-cited purposes of public attributions [of cyberattacks] is macro-level deterrence. The idea is that public naming-and-shaming of state-sponsored actors will cause the named states (and potentially other states that might be watching) to refrain from future attacks. For example, in announcing an indictment of Iranian hackers for [Distributed Denial of Service ("DDOS")] attacks on U.S. financial institutions, then-FBI Director James Comey explained, "By calling out the individuals and nations who use cyber attacks to threaten American enterprise, as we have done in this indictment, we will change behavior." U.S. officials made similar claims about the cyber sanctions executive order. In announcing the new sanctions regime, the Obama Administration's Cybersecurity Coordinator, Michael Daniel called it "a new way of both deterring and imposing costs on malicious cyber actors wherever they may be." (Eichensehr, 2020, p. 552)

Eichensehr notes that: "After the first U.S. attribution-by-indictment – the charges against Chinese [People's Liberation Army] officers for intellectual property theft – sources indicated that the Chinese military substantially scaled down its economic espionage activities. But at the same time, [Eichensehr admits] state-sponsored hacks of many kinds have continued after indictments" (Eichensehr, 2020, p. 553). Eichensehr also discusses what she calls "micro-level deterrence" against particular individuals who are deterred from future violations through indictment or the imposition of sanctions (Eichensehr, 2020, pp. 554–555). Certainly, the potential for deterrence is maximized through the use of international criminal law, which has the potential to contain far more stringent sanctions than simply "naming and shaming" – that is, simply publicly attributing the source of the cyberattack.

That said, as mentioned, the ICC cannot play a role in deterring cyberattacks unless actors (both state and nonstate actors) *realize* that certain cyberattacks, even if only a limited subset of them, *could* constitute Rome Statute (or other) crimes. In this respect, one welcome initiative is the convening of the "Council of Advisers on the Application of the Rome Statute to Cyberwarfare," a group of expert participants convened by the Permanent Mission of Liechtenstein to the United Nations and co-organized by Argentina, Austria, Belgium, Costa Rica, the Czech Republic, Estonia, Luxembourg, Portugal, Spain, and Switzerland, as well as the Global Institute for the Prevention of Aggression ("Council of Advisers," 2021). The goal of the group is to increase awareness of the potential for the Rome Statute to cover certain cyberattacks through its meetings and the eventual release of a report (see also Digital Watch discussing the Open Ended Working Group on Cybersecurity at the UN). (The author serves on the Council of Advisers.)

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

It is not claimed that this increased knowledge will fully deter cyberattacks that could be encompassed by the Rome Statute; in particular, one would expect less deterrence in situations where no ICC jurisdiction exists, and where one would not anticipate the Security Council referring a situation to the ICC (see the Rome Statute, 1998, Arts. 12(2)(a)–(b), 13(b), 15*bis*, *15ter* on jurisdiction).[3] For example, it would be naïve to anticipate referral by the Security Council of a situation to the ICC (which is permitted, Rome Statute, 1998, Arts. 13(b), 15*ter*), if a permanent member of the Security Council is involved in a cyberattack. (The permanent members hold veto power over substantive Security Council votes, UN Charter, Art. 27(3)). Additionally, it might be difficult to deter informal or rogue bands of hackers who might remain unaware of any expert report on cyberattacks (or even the ICC's existence), and perhaps would not be deterred regardless. An additional argument could be made that the ICC would have to become a more effective institution before it creates significant deterrence – for example, it has a significant number of outstanding arrest warrants (see ICC Warrant/Summonses, n.d.). Furthermore, that the ICC tends to focus its prosecutions on higher-level perpetrators further suggests that "ordinary hackers" would not necessarily fall within its focus absent an egregious cyberattack, and so decreases any deterrence potential to "ordinary hackers." Yet, the ICC is not limited to prosecuting only those bearing the "greatest responsibility" for statutory crimes, as was, for example, the Special Court for Sierra Leone (Special Court Statute, Art. 1.1); thus, if a particularly egregious cyberattack were to occur, an "ordinary hacker" could potentially attract the ICC Prosecutor's focus, including, potentially, all who aided and abetted the crime or who acted with the "common purpose" of committing the crime.[4] Notwithstanding, as mentioned, the initial first step in attempting to maximize deterrence – and thereby potentially contributing to the goal of achieving a state of cyber peace – is most certainly to create broader awareness of the ICC's potential to prosecute a limited subset of cyberattacks.[5]

The section below briefly considers two initial overarching considerations that restrict the cyberattacks the ICC might be able to prosecute. The following

[3] As to the crimes of genocide, war crimes, and crimes against humanity, the ICC has jurisdiction over crimes committed: (1) in the territory of Rome Statute States Parties; (2) by the nationals of Rome Statute States Parties, or; (3) within situations referred by the Security Council (Rome Statute, 1998, Arts. 12(2)(a)–(b); 13(b)). A state may also accept jurisdiction by entering a declaration pursuant to Article 12(3). There is a different and more restrictive jurisdictional regime for the crime of aggression, including that there is *no* jurisdiction over crimes committed in the territory of, or by the nationals of, non-States Parties (Rome Statute, 1998 Art. 15*bis*, para. 5). Referrals by the Security Council are also permitted covering the crime of aggression (Rome Statute, 1998, Art. 15*ter*).

[4] For background on individual criminal responsibility, including "aiding and abetting" and the "common purpose" doctrine, see Ambos, 2016b.

[5] For another analysis of how cyberattacks could fall within the ICC's definitions of war crimes and crimes against humanity, but finding it difficult to envision them constituting the crime of aggression, see Ambos, 2015.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

section provides a brief overview of how certain cyberattacks could fall within the Rome Statute's substantive crimes – war crimes, crimes against humanity, genocide, and the crime of aggression. A more expansive discussion of both topics can be found in my forthcoming article "The Criminalization of Cyberattacks under the International Criminal Court's Rome Statute" (Trahan, forthcoming) and the upcoming Report of the Council of Advisers on the Application of the Rome Statute to Cyber Warfare (forthcoming).

4 OVERARCHING CONSIDERATIONS AS TO ICC PROSECUTIONS

Some of the limiting factors in terms of prosecuting cyberattacks before the ICC include (1) the Rome Statute's "gravity" threshold and (2) the need to prove attribution through admissible evidence that could satisfy the standard of proof beyond a reasonable doubt. While they are beyond the scope of the present chapter, additional limiting factors include the need to satisfy jurisdiction; the ICC's "intent" requirement (which excludes responsibility for unforeseen consequences and severely restricts it even as to foreseeable consequences);[6] and the prohibition in the Rome Statute on expanding definitions of crimes by analogy, with ambiguity construed to favor the defense (Rome Statute, 1998, Art. 22(2)). (For a discussion of all three topics, see Trahan, forthcoming.)

5 THE ICC'S GRAVITY THRESHOLD

For a case to be "admissible" before the ICC, Article 17 of the Rome Statute requires that it be of "sufficiently gravity to justify … action by the Court" (Rome Statute, 1998, Art. 17(1)(d)). Article 53 further states that the Prosecutor may only initiate an investigation or proceed with a case if it "would be admissible under Article 17" (Rome Statute, 1998, Arts. 53(1)(b), 53(2)(b)). These provisions raise the question of which cyberattacks would be considered more grave and which less grave, or of marginal gravity. The ICC's cases to date have focused on rather large-scale crime scenes, with the "smaller" crime scenes probably being the killing of twelve peacekeepers, at issue in the *Abu Garda* case (see Whiting, 2015),[7] and the destruction of nine mausoleums and one mosque at issue in the *Al Mahdi* case (*Prosecutor* v. *Al Mahdi*, Case Information

[6] "The *Lubanga* appeal judgment confirmed the interpretation put forward in the *Bemba* decision on the confirmation of charges, that under Art. 30 … 'the standard for the foreseeability of events is virtual certainty.'" (Badar & Porro, 2017, Art. 30(2)(b), *citing Prosecutor* v. *Lubanga*, 2014, ICC A. Ch., "Judgment on the Appeal of Mr. Thomas Lubanga Dyilo against his conviction," paras. 441 *et seq.*; *Prosecutor* v. *Bemba*, 2009, ICC PT. Ch., "Decision Pursuant to Article 61(7)(a) and (b)," paras. 359 *et seq.*) If the ICC remains consistent with this approach, it would mean that criminal responsibility for unforeseeable consequences would be excluded and even for foreseeable consequences, the standard would be "virtual certainty" that the consequences will result.

[7] The author in no way means to minimize the severity of these crimes.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Sheet, 2018). Both cases involved the killing of persons or the destruction of physical objects.[8]

In terms of evaluating the gravity of cyberattacks, a useful starting point for analysis is *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Tallinn 2.0). Admittedly, there are divergent views among states and experts as to the weight to accord the Tallinn Manual (Efrony & Shany, 2018), and in any event they are not binding on the ICC, yet they can at least provide a useful starting point.

In Tallinn 2.0, the experts focused on what constitutes an "armed attack" committed through cyber means. They engaged in this analysis because an "armed attack" can justify self-defense under Article 51 of the UN Charter (see UN Charter, Art. 51); they were not engaging in this analysis in relationship to the ICC. Tallinn 2.0 takes the position that "a cyber operation that seriously injures or kills a number of persons or that causes significant damage to, or destruction of, property would satisfy the scale and effects requirement" and thus constitute an "armed attack" (Schmitt, 2017, Rule 71, para. 8.) This would provide one possible standard by which to evaluate the gravity of cyberattacks. Namely, only if a cyberattack seriously injures or kills a number of persons or causes significant damage to, or destruction of, property, would that satisfy Rome Statute gravity.

The experts additionally suggested various criteria that "States are likely to consider" as to when a cyber operation constitutes a "use of force" (relevant to considering when Article 2(4) of the UN Charter is violated, see UN Charter, Art. 2(4)). These criteria included severity, immediacy, directness, invasiveness, and measurability of effects (Schmitt, 2017, Rule 69, para. 9 (a)–(e)). These would appear useful criteria to consider in making the gravity evaluation. Additional factors that could prove useful for consideration include those identified in ICC case law (see, e.g., *Prosecutor* v. *Al Hassan*, 2020, paras. 59, 89, 90) and by the Prosecutor (see "Policy Paper on Preliminary Examinations," 2013, paras. 62, 64–65).

Another interesting consideration is whether loss of life or physical destruction is always the most grave of harms? For instance, France takes the position that operations that penetrate military systems to weaken French defensive capabilities, even if this does not produce physical effects, would constitute a "use of force" (Droit International, 2019, as cited in Schmitt, 2019).[9] Similarly, the Netherlands takes the view that a catastrophic systems attack that causes very serious economic impact could constitute an "armed attack" (Schmitt, 2019, quoting the Dutch Minister of Defence). While neither country is opining on whether such conduct would meet Rome Statute gravity, it is worth considering whether it should do so. The author suggests that here one might differentiate between penetration of military systems and catastrophic systems attacks that cause serious

[8] For a general discussion of gravity, see deGuzman, 2020. See also n. 10 (discussing the OTP's not proceeding in the *Comoros* case where there were ten fatalities).

[9] For France's most significant statement regarding the application of international law in cyberspace, see Droit International, 2019, as cited in Schmitt, 2019.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

injuries or fatalities to persons, or significant damage to or destruction of property, from those that do not. Another interesting question is whether the destruction of "data" should be encompassed, or whether the "property" destroyed would need to be physical property (see, e.g., Biller and Schmitt, 2019; Mačák, 2015; Horowitz, 2020, considering destruction of data).

The ICC might wish to examine these issues and consider issuing a policy paper on application of the Rome Statute to cyberattacks (including the topic of gravity). At the same time, in terms of increasing deterrence potential, there could also be some advantages flowing from ambiguity. Thus, this author suggests that one possible gravity threshold for ICC prosecution could be where a cyberattack causes serious injuries or fatalities to persons or significant damage to or destruction of property; at the same time, perhaps one would not want to close the door to other large-scale or invasive attacks that do not meet this definition. Also, as the ICC Appeals Chamber explains in the *Al Hassan* case, the gravity requirement does not "oblige the Court to choose only the most serious cases, but merely [obliges] it not to prosecute cases of marginal gravity" (*Prosecutor* v. *Al Hassan*, 2020, para. 59).

The ICC's pursuing of a case where a cyberattack of sufficient gravity occurs would be significant in itself, and make clear that the Rome Statute *does* encompass cyberattacks. Roscini writes: "the Prosecutor might decide to select certain situations and cases involving the commission, instigation, or facilitation of international crimes through cyber conduct because of their impact or to deter them in the future, even if they resulted in a lower number of victims than in other cases" (Roscini, 2019, p. 271).[10] This "expressivist" approach – pursuing prosecutions that further protected values and thereby sending a message to achieve a given result – is indeed an important and legitimate aspect of prosecutorial strategy (Cross, 2020, pp. 67–68).

6 PROVING ATTRIBUTION THROUGH ADMISSIBLE EVIDENCE THAT ESTABLISHES PROOF BEYOND A REASONABLE DOUBT

An additional limiting factor – true for all ICC crimes – would be that all the elements of the crime would need to be proven through admissible evidence that could eventually satisfy the requirement at trial of proof beyond a reasonable doubt (Rome Statute, 1998, Art. 66(3)). This includes the issue of attribution (who conducted the cyberattack), sourcing it not just to a state (or nonstate actor/hacker group working for the state) but potentially to a particular "computer, … to identify the person who operated the computer, and more importantly to identify the real 'mastermind'

[10] In the *Comoros* case, the ICC's Office of the Prosecutor (OTP) took the position that ten fatalities did not meet the gravity threshold, although there were other considerations than simply the number of fatalities (OTP, Report on Preliminary Examination Activities, 2017, para. 336). In the *Abu Garda* case, as mentioned, the OTP did proceed regarding the killing of twelve peacekeepers "because of the significance of the target and impact on peacekeeping operations" (Whiting, 2015).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

behind the attack …." (Tsagourias, 2012, p. 233). This could pose significant challenges (Dederer & Singer, 2019, p. 438 (citing sources)).

Compounding difficulties, cyberattackers sometimes go to lengths to conceal cyber operations (Hathaway et al., 2012, p. 843), and states sometimes deliberately hide their attack as perpetrated by another state ("false-flagging"). For example, this happened when "Russian hackers piggy-backed on an Iranian cyber-espionage operation," thereby hacking into "government and industry organizations in dozens of countries while masquerading as attackers from the Islamic Republic [of Iran]" (Stubbs & Bing, 2019). Or states can hide behind nonstate actors to mask their operations (Dederer & Singer, 2019, p. 438; Hathaway et al., 2012, p. 854). Even when that is not the case, because cyberattacks can be perpetrated through a single computer or network of computers located far from where the consequences impact, they can be extremely difficult to attribute (Dederer & Singer, 2019, p. 431; Brenner, 2011, p. 32). The attacks can also be concealed by feigning that operating systems are functioning normally (Rowe, 2007; Hathaway et al., 2012, p. 828).

Furthermore, if attribution is to be made, this raises questions as to who would be in a position to do so. Would this be practical for the ICC to do itself? And, if not, what are the implications of relying on state cooperation in this regard? In addition to attribution, all of the evidence in the case would require "authentication," and this (and simply having the knowledge to assemble a cyberattack case) would require significant technical expertise. Relying on state cooperation also carries pitfalls in that states may be more likely to cooperate when it suits their self-interests (e.g., they have suffered from a cyberattack), and not cooperate when it does not serve their interests (e.g., they were the perpetrator or linked to the perpetrator). Thus, there will be significant challenges in terms of attribution, authentication, and development of the necessary expertise to establish both. Building ICC expertise will require both the hiring of staff, and/or use of outside experts, and development of relevant policies.

Thus, the above discussion suggests the potential applicability of the Rome Statute to a limited subset of cyberattacks: (1) if they meet the Rome Statute's gravity threshold and (2) where attribution could be proven beyond a reasonable doubt. As mentioned, other limiting factors include whether jurisdiction exists; whether the "intent requirement" can be proven (which appears to exclude responsibility for unforeseen consequences and limit responsibility even for foreseen consequences);[11] and whether the crimes can be applied without drawing on analogies, with ambiguity construed to favor the defense (Trahan, forthcoming).

7 THE ROME STATUTE'S SUBSTANTIVE CRIMES

Despite the limitations suggested above, the next section outlines the key elements of the Rome Statute's four core crimes – genocide, war crimes, crimes

[11] See supra note 7.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

against humanity, and the crime of aggression – and suggests how a limited subset of cyberattacks might fall within the definitions of each.[12] Again, domestic jurisdictions, even ones that incorporate these crimes into their domestic criminal codes, could adopt broader definitions of the crimes; thus, the elements of the crimes discussed below would not necessarily apply in domestic jurisdictions, which also might have or develop broader criminal statutes covering cyberattacks and/or cybercrimes.

8 CYBERATTACKS AS WAR CRIMES UNDER THE ROME STATUTE

As mentioned, the rules of international humanitarian law apply in the cyber domain (Schmitt, 2017, Rule 80; "The Paris Call," 2018). Thus, for example, Tallinn 2.0 explains that in a state of armed conflict, cyberattacks may not target civilians (Schmitt, 2017, Rule 80), may not be indiscriminate (Schmitt, 2017, Rule 105), and may not cause excessive "collateral damage" (Schmitt, 2017, Rule 113). Tallinn 2.0 expressly acknowledges that when such IHL rules are violated, "[c]yber operations may amount to war crimes and thus give rise to individual criminal responsibility under international law" (Schmitt, 2017, Rule 84).

Under the Rome Statute, "[the] Court shall have jurisdiction in respect of war crimes in particular when committed as a part of a plan or policy or as part of a large-scale commission of such crimes" (Rome Statute, 1998, Art. 8(1)). Additionally, all of the contextual elements for war crimes would need to be proven – such as the existence of an "armed conflict" (whether international or noninternational), a "nexus" between the cyberattack and the armed conflict (ICC, "Elements of Crimes," 2011), and that the perpetrator was aware of the factual circumstances that established the existence of the armed conflict (ICC, "Elements of Crimes," 2011). There would also be the elements for the specific underlying war crime(s), as well as – as explained above – the need to prove attribution (linking a specific perpetrator), intent, and jurisdiction. As to specific war crimes, note that the Rome Statute contains different lists of war crimes depending on whether the crimes were committed during international armed conflict or noninternational armed conflict (*compare* Rome Statute, 1998, Art. 8.2(a)–(b) *with* Art. 8.2(c), (e)).

As to the requirement of armed conflict, under the generally accepted definition from the ICTY's *Tadić* case, "an armed conflict exists whenever there is a resort to armed force between States or protracted armed violence between governmental authorities and organized armed groups or between such groups within a State" (*Prosecutor* v. *Tadić*, 1995, para. 70). The Rome Statute (and IHL) particularly

[12] The implications of the author's argument – that certain cyberattacks fall within the Rome Statute's existing crimes – suggests that in terms of retroactivity, jurisdiction for qualifying cyberattacks would be the same as it is for the crimes generally. That is, for initial ratifying States Parties, it could go back to July 1, 2002, and for the crime of aggression it could go back to July 17, 2018.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

exclude "situations of internal disturbances and tensions, such as riots, isolated and sporadic acts of violence or other acts of a similar nature" (Rome Statute, 1998, Art. 8(2)(d)). An international armed conflict is one in which two or more states are parties to the conflict (Rona, 2003, p. 58; Common Article 2 to the 1949 Geneva Conventions). Noninternational armed conflict is defined as "armed conflict not of an international character" (Common Article 3 to the 1949 Geneva Conventions). For noninternational armed conflict, the operations must also have reached a minimum level of intensity and a nonstate armed group must have achieved a certain level of organization (*Prosecutor* v. *Tadić*, 1995, para. 70).[13]

The cyberattacks on Eastern Ukraine provide a possible example of war crimes perpetrated using, at least partly, cyberattacks. The attacks there were "blended attacks," perpetrated through cyber and physical/kinetic means. In addition to the armed conflict that killed more than 10,000, the hacking into dozens of governmental organizations and companies through a "scorched-earth" cyberattack, which penetrated victims ranging from media outlets to railway firms and hospitals, caused hundreds of thousands of homes to lose electricity and shut down at least three regional utilities (Greenberg, 2018; Bezhan, 2016; Efrony & Shany, 2018, pp. 624–626). Both the United States and the United Kingdom believe that the cyberattacks on Ukraine were perpetrated by Russia's military intelligence service, commonly known by the Russian acronym GRU (Warrell et al., 2020).

The cyberattack, conducted during a state of armed conflict could, if all the elements of the crimes were able to be proved through admissible evidence, potentially constitute the war crime of intentionally directing attacks against the civilian population (Rome Statute, 1998, Art. 8(2)(b)(i)), or civilian objects (Rome Statute, 1998, Art. 8(2)(b)(ii)),[14] or inflicting "collateral damage" – incidental loss of life or injury to civilians that is "clearly excessive in relationship to the concrete and direct overall military advantage anticipated" (Rome Statute, 1998, Art. 8(2)(b)(iv)).[15] The cyber operations also appear to have been "indiscriminate."[16] A cyberattack against a medical facility – of which there were several in Eastern Ukraine (Greenberg, 2018) – could also constitute a war crime under Rome Statute Articles 8(2)(b) (xxiv) and

[13] For one analysis of when a cyberattack reaches the threshold of armed conflict, see Ambos, 2015, at pp. 121–126. Ambos also notes that groups of "hackers" may not meet the organization requirement (Ambos, 2015, at pp. 125, 129).

[14] Application of the principle of distinction may, however, be complicated by "the *interconnectivity* between military and civilian computer systems and the mostly *dual-use* of cyber ingrastructure," although "dual-use objects are qualified as military objectives since they normally contribute to military purposes . . . ." (Ambos, 2015, at p. 131) (italics in original).

[15] For analysis of when a civilian "directly participates" in hostilities in the cyber context, so as to become a permissible target, see Ambos, 2016a, at p. 128.

[16] Here, the Rome Statute has a problem. Rome Statute Article 8(2)(b)(xx) prohibits employing weapons that are "inherently indiscriminate," but only for weapons "included in an annex to th[e] Statute"; puzzlingly, there is no such annex (see Clark, 2009). Thus, at present, use of inherently indiscriminate weapons cannot be prosecuted at the ICC (unless their use also happens to constitute another war crime).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

(e)(ii) (Mačák et al., 2020). As with all ICC crimes, one would, among other things, additionally need to attribute responsibility to particular individuals for an ICC case to proceed and satisfy the intent requirement, both of which could prove difficult. There is ICC jurisdiction over the events in Ukraine because Ukraine executed an Article 12(3) declaration, accepting the ICC's jurisdiction over crimes committed on its territory from November 21, 2013 to February 22, 2014, and then executed another such declaration covering crimes committed from February 22, 2014 and continuing on an open ended basis. (ICC Investigation, Ukraine, n.d.). Thus, there currently is ICC jurisdiction over cyberattacks that have been and are being committed in Ukraine, as well as jurisdiction over war crimes, crimes against humanity, and genocide more generally.

9 CYBERATTACKS AS CRIMES AGAINST HUMANITY UNDER THE ROME STATUTE

Crimes against humanity are defined in the Rome Statute as acts "committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack" (Rome Statute, 1998, Art. 7(1)). The "attack" against the civilian population is defined as "a course of conduct involving the multiple commission of acts [enumerated in Article 7(1)] against any civilian population, pursuant to or in furtherance of a State or organizational policy to commit such attack" (Rome Statute, 1998, Art. 7(2)(a)).[17] For crimes against humanity, the attack is directed against a civilian population and need not be a military attack or linked to armed conflict (see, e.g., *Prosecutor* v. *Ntaganda*, 2019, para. 662). There are also requirements that the perpetrator's "conduct was committed as part of a widespread or systematic attack directed against a civilian population" (the "nexus" requirement) and that "[t]he perpetrator knew that the conduct was part of or intended the conduct to be part of a widespread or systematic attack directed against a civilian population" (ICC, "Elements of Crimes," 2011). The "underlying crimes" that support a charge (or multiple charges) of crimes against humanity are murder, extermination, enslavement, deportation, imprisonment, torture, rape or sexual violence, persecution, enforced disappearances, apartheid, and other inhumane acts (see Rome Statute, 1998, Art. 7(1)(a)–(k) for details).

Let us assume a "cyber-9/11" scenario, where the attackers have used cyber means to jam the controls of several airplanes, causing them to crash into buildings with ensuing large-scale loss of life. That would likely constitute the crime against humanity of murder if evidence proves that the attack was "widespread" (e.g., impacting a large number of victims) or "systematic" (a coordinated, organized attack) and orchestrated through a "State or organizational policy" (proof of which may be inferred,

[17] See Ambos, 2015, at p. 142 ("While a loosely organized group of hackers acting autonomously would not meet the organization requirement, organized armed groups within the meaning of IHL that take recourse to methods of cyber warfare certainly would.").

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Prosecutor v. *Bemba*, 2016, para. 160), and one can attribute responsibility to particular perpetrators, prove intent, and satisfy jurisdictional requirements. The same cyberattack, if directed toward members of a particular protected group, could additionally constitute the crime against humanity of persecution (see Rome Statute, 1998, Art. 7(1)(h), listing protected groups). Crimes against humanity also include a residual "catch-all" – namely, "[o]ther inhumane acts of a similar character [to other crimes against humanity] intentionally causing great suffering, or serious injury to body or to mental or physical health" (Rome Statute, 1998, Art. 7(1)(k). Cyberattacks with severe consequences, such as a cyber 9/11, could also fall within this category.

While there appears to be great interest and concern about the problem of cyberattacks disrupting elections, to this author such interference – which could certainly be "widespread" *and* "systematic" (although it need not be both) – does not rise to the level of "other inhumane acts" because it would not involve "great suffering, or serious injury to body or to mental or physical health." It also does not appear to fit into any of the other "underlying crimes" of crimes against humanity (see Rome Statute, 1998, Art. 7(1)(a)–(k)).[18]

10 CYBERATTACKS AS GENOCIDE UNDER THE ROME STATUTE

Genocide is a crime that targets members of a distinct "national, ethnical, racial or religious group" (Rome Statute, 1998, Art. 6). For this crime, it is not the attack itself, but the intent behind the attack that is key. The *dolus specialis* (special mental state requirement) of genocide requires proof of: (1) "intent to destroy"; (2) "in whole or in part"; (3) of a "national ethnical, racial or religious group"; and (4) "as such" (i.e., because individuals belong to such a group) (*ibid.*; Kreβ, 2006, p. 498). While genocide includes "inchoate" forms – for example, incitement to commit genocide could be the completed crime (Rome Statute, 1998, Art. 25(3)(e); Ohlin, 2009, discussing "inchoate crimes") – for Rome Statute purposes, if no genocide occurs the crimes probably would not satisfy ICC gravity requirements.

[18] Although some might argue that it could constitute "persecution" against the nationals of another country, that would certainly involve a novel reading of what constitutes persecution, and any ambiguity in Rome Statute crimes, as explained above, must be construed to favor the defense (Rome Statute, 1998, Art. 22(2)). If one does not have an "underlying crime," then pursuant to Article 7(2)(a), the "attack" requirement for crimes against humanity also is not met. Note additionally that as to Russian interference in US elections (see, e.g., Lewis, 2020; Ohlin, 2020), because that involves the territories and nationals of two non-States Parties, there would also be no ICC jurisdiction (see Rome Statute, 1998, Art. 12(2)(a)–(b)), unless, for example, the United States were to enter an Article 12(3) declaration accepting ICC jurisdiction Rome Statute, 1998, Art. 12(3)) – a rather unlikely scenario. By contrast, election interference in various European states (which also has occurred, Apuzzo & Satariano, 2019) who are ICC States Parties would be within ICC jurisdiction as long as an element of the crime occurred in the territory of a State Party (Myanmar/Bangladesh decision, 2019); yet, that is probably moot because this author does not view election interference as meeting the requirements of crimes against humanity (nor any other Rome Statute crime). See discussion below analyzing election interference as a crime of aggression – but concluding it likely also does not meet that definition.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

In addition to these overall requirements, there must be "underlying crimes"; the first enumerated being the killing of members of a group (Rome Statute, 1998, Art. 6(a)). The second underlying crime is "[c]ausing serious bodily or mental harm to members of the group" (Rome Statute, 1998, Art. 6(b)). The third underlying crime is "[d]eliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part" (Rome Statute, 1998, Art. 6(c)) – see also Arts. 6(d)–(e)). Again, all are subject to Rome Statute gravity requirements, as one can also imagine a "mental" harm caused by a cyberattack that does not rise to the level of Rome Statute gravity, or creating horrible conditions of life for members of a protected group that is not necessarily aimed at bringing about the group's physical destruction, and/or does not meet Rome Statute gravity. Thus, for a cyberattack to constitute the crime of genocide, it would need to satisfy both this overall special mental state requirement and proof of at least one of the underlying crimes. Additionally, as with all Rome Statute crimes, proof of attribution to particular individuals, proof of intent, and jurisdiction are required.

Here, it may be easier to envision cyber enabled genocide. In Rwanda, in 1994, *Radio Télévision Libre des Mille Collines* (RTLM) was used to incite and facilitate the killing of members of the Tutsi ethnic group – with the Tutsi identified by their government – issued identity cards, particularly at roadblocks (Metzi, 1997). One can similarly imagine cyber means used to compromise hospital or other medical records to identify members of a protected group, and/or; cyber means being used to incite genocide against protected group members (see, e.g., Mozur, 2018, discussing Burmese military Facebook incitement, coupled with crimes against the Rohingya). In either situation, assume the identification of protected group members and/or incitement is followed by killings (as it was in Rwanda and Myanmar), and one could infer the required genocidal intent (see, e.g., *Prosecutor* v. *Akayesu*, 1998); Burmese Military Document entitled "Rohingya Extermination Plan," Mansour, 2017). Either could satisfy the elements of genocide.[19] Roscini also provides the example of a cyberattack that shuts down the cooling system of a nuclear power reactor releasing high levels of radiation killing members of a particular national group, if one could prove genocidal intent (Roscini, 2019, p. 250).

11 CYBERATTACKS AS THE CRIME OF AGGRESSION UNDER THE ROME STATUTE

While the crime of aggression has numerous requirements and warrants a far more extensive discussion (see Trahan, forthcoming), some of the key requirements are that there is a state "act of aggression" (Rome Statute, 1998, Art. 8*bis*, para. 2) that,

[19] The ICC has limited jurisdiction related to crimes against the Rohingya. It only has jurisdiction where one element of the crime occurred in the territory of a Rome Statute State Party (Bangladesh), but not as to crimes committed solely within Myanmar.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

to qualify as the crime of aggression, must also constitute a "manifest" violation[20] of the UN Charter by its "character, gravity and scale" (Rome Statute, 1998, Art. 8*bis*, para. 1). The "act of aggression" is defined as "use of armed force by a State against the sovereignty, territorial integrity or political independence of another State, or in any other manner inconsistent with the Charter of the United Nations"[21] (Rome Statute, 1998, Art. 8*bis*, para. 2). There is also a list of acts enumerated in Rome Statute Article 8*bis*, paragraph 2 (a)–(g) that could meet that requirement, but each would additionally need to constitute a "manifest" violation of the UN Charter (Rome Statute, 1998, Art. 8*bis*, para. 2 (a)–(g)).[22] Another requirement is that the crime only covers "leaders" in that it applies only to "person[s] in a position effectively to exercise control over or to direct the political or military actions of a State"[23] (Rome Statute, 1998, Art. 8*bis*, para. 1). Also, the leader would need to engage in the "planning, preparation, initiation or execution" of the crime (Rome Statute, 1998, Art. 8*bis*, para. 1).

While the above requirements appear difficult to satisfy, the fourth act enumerated as potentially qualifying as an "act of aggression" is "[a]n attack by the armed forces of a State on the land, sea or air forces, or marine and air fleets of another State" (Rome Statute, 1998, Art. 8*bis*, para. 2(d)) and Article 8*bis* refers to "the use of *any weapon* by a State against the territory of another State" (Rome Statute, 1998, Art. 8*bis*, para. 2(b) (emphasis added)). Imagine a cyber unit within the armed forces of a state causes weapon systems of the armed forces of another state to become completely inoperable.[24] That would constitute an "attack by the armed forces of a State" on the forces of another state. One would additionally need to attribute responsibility to a particular state leader (or head of cyber command) of the attacking armed forces who is "in a position effectively to exercise control over or to direct the political or military actions of a State," and who planned, prepared, initiated, or played a role in the execution of the cyberattack. Changing the scenario slightly, imagine the leader or head of cyber command instead employs bands of nonstate hackers to conduct the same attack and later acknowledges those acts as acts of the state. Here one would look to the rules

[20] The purpose of the "manifest" requirement is "to exclude minor incidents (e.g., border skirmishes) or legally controversial cases (e.g., a humanitarian intervention) .... (Ambos, 2015, at p. 140).

[21] For analysis of how a cyberattack could constitute the use of "armed force," see Ambos, 2015, at pp. 138–139.

[22] The list of acts of aggression is "open-ended" in that Article 8*bis*, paragraph 2 lists acts that "shall" qualify as acts of aggression, leaving open that other acts might as well. Yet, charging acts not listed might prove risky, as it could run afoul of the principle *nullum crimen sine lege* (no crime without law) and the requirement that ambiguity in the Rome Statute favors the defense (Rome Statute, 1998, Art. 22(2).

[23] For further analysis of the "leadership clause" and its application regarding cyber operations, see Ambos, 2016a.

[24] "[A] cyber operation leaving the targeted object physically intact but neutralizing it in its functionality may amount to a militarily relevant attack." (Ambos, 2015, p. 124, writing this in the context of war crimes).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

on state attribution to determine whether the acts of the nonstate actors become attributable to the state, with perhaps the clearest situation being where the hackers are hired into the state cyber command structure, so they become part of the armed forces.[25] In any event, the cyberattack would need to be "manifest," such that it is not *de minimis* (insufficient in gravity and/or scale) and/or "super clear" in terms of its illegality (meeting the required "character"). Yet, because of seemingly extensive jurisdictional limitations – if they are valid (see Trahan, 2018) – it could be quite difficult to trigger ICC jurisdiction regarding the crime of aggression, absent a UN Security Council referral.[26]

Returning to the example of election interference, this author doubts that the elements of the crime of aggression would be satisfied. While election interference could be viewed as a "sovereignty violation" (Shackelford, 2017, 11; Efrony & Shany, 2018, 640), this author does not see it rising to the level of a "manifest" Charter violation, which is required for the crime of aggression (Rome Statute, 1998, Art. 8*bis*, para. 1). Furthermore, at least one significant state involved in election interference, the Russian Federation (Ohlin, 2020), is not a party to the Rome Statute, so there would be no ICC jurisdiction over the crime of aggression committed by Russian nationals (*ibid.*, Art. 15*bis*, para. 5).[27]

12 CONCLUSION

This chapter has briefly touched on what will need to be a far more extensive study considering how the crimes in the ICC's Rome Statute could potentially encompass certain cyberattacks. Yet, hopefully, this chapter has made the case that there is at least some potential for applicability. My forthcoming article and the forthcoming report of the Council of Advisers on the Application of the Rome Statute to Cyberwarfare will expand significantly on these topics.

It is important to engage in this analysis, as there would need to be broad recognition of the ICC's ability to prosecute certain cyberattacks if there is to be any

[25] The International Law Commission in its Articles on State Responsibility for Internationally Wrongful Acts discusses when acts by nonstate actors are attributable to a state (see ILC Articles, 2001, Arts. 5, 8, 9, 11; see also Efrony & Shany, 2018, p. 584; Schmitt 2017, Tallinn 2.0, Rule 14, on attribution).

[26] Absent a UN Security Council referral, if the restrictive interpretation in a certain 2017 Activating Resolution is upheld (ICC, 2017), the ICC would only have jurisdiction over the crime where a State Party that has ratified the Kampala amendment attacks another State Party that has also ratified the Kampala amendment (see Trahan, 2018).

[27] Ironically, this was an exemption that the US delegation negotiated, possibly supported by a few other states, at the ICC Review Conference in Kampala, Uganda (Trahan, 2011). Here, unlike with crimes against humanity, election interference by a non-State Party *even against Rome Statute State Parties* would fall outside ICC jurisdiction (Rome Statute, 1998, Art. 15*bis*, para. 5). Furthermore, due to the veto power of the permanent members of the UN Security Council (UN Charter, Art. 27(3)), one can also anticipate there would be no Security Council referral.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

potential for deterrence. Only then can international criminal law in this area play a role in maximizing the potential of reaching a state of cyber peace. It is actually quite significant that there is an existing international criminal tribunal with jurisdiction to prosecute a limited subset of cyberattacks. This capacity was probably never envisioned when the Rome Statute was negotiated; yet, certain cyberattacks appear to meet the elements of the ICC crimes. Whether it is feasible to bring cases will depend if attribution can also be established, and if all of the elements of the crime can be proven through admissible evidence that satisfies the standard, at trial, of proof beyond a reasonable doubt. Perhaps this is not fully feasible now, but as technology develops, it could become more achievable in the future.

None of the cyberattacks perpetrated to date probably have reached the threshold for Rome Statute crimes with the possible exception of those in Ukraine, over which the ICC has an open preliminary examination (ICC, "Preliminary Examination, Ukraine," n.d.). It may also take time for the ICC to develop the required expertise to be able to develop and prosecute such cases, and the ICC may need to rely extensively on the outside expertise of cyber experts. Yet, as mentioned, that also carries potential pitfalls. To the extent the ICC can develop its own internal capacity that could help alleviate potential conflicts of interest.

International criminal law *does* have a role to play here. Will this deter all cyberattackers from committing grievous cyberattacks? The author will not be so naïve to claim that it will. But if the ICC is able to achieve some deterrence – deterring even one horrific cyberattack – that would certainly be a worthwhile endeavor. Ironically, it will be hard to know if such an attack has been deterred, because it would involve the absence of the attack, something notoriously difficult to prove.

While ICC States Parties may be "willing" and "able" to prosecute cyberattacks, and under Article 17 of the Rome Statute, that would then render a case "inadmissible" before the ICC (see Rome Statute, 1998, Art. 17), it is also quite possible that domestic jurisdictions will lack the required laws and/or be unable to exercise jurisdiction over the totality of the crime (which potentially might involve a foreign attacker state and multiple "victim" states). Then, the domestic jurisdiction would be "unable" to prosecute the case fully, likely rendering the case "admissible" before the ICC if other Rome Statute requirements are also satisfied.

To date, most of the ICC's focus has been on crimes in developing countries. Because both developed and developing countries suffer from cyberattacks (probably developed countries even more so), a focus on such crimes before the ICC could be a welcome development, at least in the eyes of many ICC States Parties. Promoting the applicability of the Rome Statute to certain cyberattacks could additionally demonstrate an increased relevance of the ICC to one of the more vexing contemporaneous challenges facing the international community.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

REFERENCES

Acquaviva, G. (2014). International criminal courts and tribunals as actors of general deterrence? Perceptions and misperceptions. *International Review of the Red Cross*, 96(895), 784.

Ambos, K. (2015). International criminal responsibility in cyberspace, in N. Tsagourias & R. Buchan (Eds), *Research handbook on cyberspace and international law*, Edward Elgar, 118.

Ambos, K. (2016a). Individual criminal responsibility for cyber aggression. *Journal of Conflict & Security Law*, *21*(3), 495.

Ambos, K. (2016b). Article 25. Individual criminal responsibility, in O. Triffterer & K. Ambos (Eds), *The Rome Statute of the International Criminal Court: A commentary* (3rd edn, C.H. Beck, Hart, Nomos, 2016), 979.

Apuzzo, M., & Satariano, A. (2019, May 12). Russia Is Targeting Europe's Elections. So Are Far-Right Copycats. *The New York Times*. Retrieved from www.nytimes.com/2019/05/12/world/europe/russian-propaganda-influence-campaign-european-elections-far-right.html

Badar, M. E., & Porro, S. (2017, August 18). Article 30(2)(b), Intent in Relation to Result. *Case Matrix Network*. Retrieved from Case Matrix Network.

Baezner, M., & Robin, P. (2018, January). Hotspot Analysis: Cyber and Information Warfare in the Ukrainian Conflict. CSS Cyber Defense Project.

Bellovin, S. M., Landau, S., & Lin, H. S. (2017). Limiting the undesired impact of cyber weapons: Technical requirements and policy implications. *Journal of Cybersecurity*, 3(1), 59.

Berger, J. (2016, March 26). A Dam, Small and Unsung, Is Caught Up in an Iranian Hacking Case. *The New York Times*. Retrieved from www.nytimes.com/2016/03/26/nyregion/rye-brook-dam-caught-in-computer-hacking-case.html

Bezhan, F. (2016, January 5). *Cyberattack on Ukrainian Power Grid Looks to Some Like an Apocalyptic First*. Radio Free Europe. Retrieved from www.rferl.org/a/ukraine-blackout-cyberattack-power-grid-apocalyptic-first/27469154.html

Biller, J. T., & Schmitt, M. N. (2019). Classification of cyber capabilities and operations as weapons, means, or methods of warfare. *International Law Studies*, 95, 179.

Brenner, J. (2011). *America the vulnerable: Inside the new threat matrix of digital espionage, crime, and warfare*. Penguin Press.

Brierly, J. L. (1944). *The outlook for international law*. Clarendon Press.

Clark, R. S. (2009). Building on article 8(2)(b)(xx) of the Rome Statute of the International Criminal Court: Weapons and methods of warfare. *New Criminal Law Review*, *12*(3), 366.

Cross, M. E. (2020). Strategising international prosecutions: How might the work of the Kosovo specialist prosecutor's office come to be judged? *International Criminal Law Review*, 20(1), 43.

Decision Pursuant to Article 15 of the Rome Statute on the Authorisation of an Investigation into the Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar, Case No. ICC-01/19-27, Judgment, ¶ 61 (November 1, 2019) ("Myanmar/Bangladesh decision, 2019").

Dederer, H-G., & Singer, T. (2019). Adverse cyber operations: Causality, attribution, evidence, and due diligence. *International Law Studies*, 95(1), 430.

deGuzman, M. M. (2020). *Shocking the conscience of humanity: Gravity and the legitimacy of international criminal law*. Oxford University Press.

Droit International Appliqué aux Opérations dans le Cyberspace. (2019). Just Security. Retrieved from www.justsecurity.org/wp-content/uploads/2019/09/droit-internat-appliqu%C3%A9-aux-op%C3%A9rations-cyberespace-france.pdf

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Efrony, D., & Shany, Y. (2018). A rule book on the shelf? Tallinn manual 2.0 on cyber operations and subsequent state practice. *American Journal of International Law*, *112*(4), 583.

Eichensehr, K. (2020). The law & politics of cyberattack attribution. *UCLA Law Review*, *67*, 520.

Geneva Convention I for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (1949, August 12). 75 UNTS 31. Geneva Convention II for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea. (1949, August 12). 75 UNTS 85. Geneva Convention III Relative to the Treatment of Prisoners of War. (1949, August 12). 75 UNTS 135. Geneva Convention IV Relative to the Protection of Civilian Persons in Time of War. (1949, August 12). 75 UNTS 287. (Collectively, "1949 Geneva Conventions").

Glaser, A. (2017, June 27). *U.S. Hospitals Have Been Hit by the Global Ransomware Attack*. Vox. Retrieved from www.vox.com/2017/6/27/15881666/global-eu-cyberattack-us-hackers-nsa-hospitals

Greenberg, A. (2018, August 22). The Untold Story of NotPetya, the Most Devastating Cyberattack in History. *Wired*. Retrieved from www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

Greenberg, A. (2017a, July 6). Hack Brief: Hackers Targeted a US Nuclear Plant (But Don't Panic Yet). *Wired*. Retrieved from www.wired.com/story/hack-brief-us-nuclear-power-breach/#intcid=recommendations_wired-bottom-recirc-similar_1691318a-b422-4428-96db-e7512a834566_text2vec1_text2VecSimilarity

Greenberg, A. (2017b, September 6). Hackers Gain Direct Access to US Power Grid Controls. *Wired*. Retrieved from www.wired.com/story/hackers-gain-switch-flipping-access-to-us-power-systems/#intcid=recommendations_wired-bottom-recirc-similar_712531e1-69ed-4994-b83b-891af026859f_text2vec1_text2VecSimilarity

Greenfield, R. (2013, April 23). Look What the Hacked AP Tweet About White House Bombs Did to the Market. *The Atlantic*. Retrieved from www.theatlantic.com/technology/archive/2013/04/hacked-ap-tweet-white-house-bombs-stock-market/315992/

Griffiths, J. (2015, October 8). Cybercrime Costs the Average U.S. Firm $15 Million a Year. *CNN Tech*. Retrieved from https://money.cnn.com/2015/10/08/technology/cybercrime-cost-business/

Hathaway, O. A., Crootof, R., Levitz, P., Nix, H., Nowlan, A., Perdue, W., & Speigel, J. (2012). The law of cyber attack. *California Law Review*, *100*(4), 817.

Hillebrecht, C. (2016). The deterrent effects of the international criminal court: Evidence from Libya. *International Interactions*, *42*(4), 616.

Horowitz, J. (2020, May 19). Cyber operations under international humanitarian law: Perspectives from the ICRC. *ASIL Insights*, *24*(11).

Human Rights Watch. (2009, July 7). *Selling Justice Short, Why Accountability Matters for Peace*. HRW. Retrieved from www.hrw.org/en/node/84262/section/2

ICC. (2011). *Elements of Crimes*. Retrieved from www.icc-cpi.int/resource-library/Documents/ElementsOfCrimesEng.pdf

ICC. (n.d.). *Preliminary Examination, Ukraine*. Retrieved from www.icc-cpi.int/ukraine#:~:text=On%2025%20April%202014%2C%20the,in%20Crimea%20and%20eastern%20Ukraine

ICC. (n.d.). *States Parties, Chronological List*. Retrieved from https://asp.icc-cpi.int/en_menus/asp/states%20parties/Pages/states%20parties%20_%20chronological%20list.aspx

ICC. (n.d.). *Warrant/Summonses*. Retrieved from www.icc-cpi.int/cases#Default=%7B%22k%22%3A%22%22%7D#2ae8b286-eb20-4b32-8076-17d2a9d9a00e=%7B%22k%22%3A%22%22%7D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

ICC Forum. (2018, February 22). New Frontiers for the ICC (International Criminal Court): Tackling Cyber Attacks through the Crime of Aggression. Retrieved from https://iccforum.com/forum/permalink/110/13832

Int'l L. Comm'n. (2001). Draft Articles on the Responsibility of States for Internationally Wrongful Acts, with Commentaries (adopted). UN Doc. A/56/10. ("ILC Articles").

International Criminal Court. (2017, December 14). Activation of the Jurisdiction of the Court Over the Crime of Aggression. Retrieved from https://asp.icc-cpi.int/iccdocs/asp_docs/Resolutions/ASP16/ICC-ASP-16-Res5-ENG.pdf

Jensen, E. T. (2017). The Tallinn manual 2.0: Highlights and insights. *Georgetown Journal of International Law*, 48, 735.

Jo, H., & Simmons, B. A. (2016). Can the International Criminal Court deter atrocity? *International Organization*, 70(3), 443.

Kampala amendment. (2010, June 2011, adopted by consensus). RC/Res.6*. Review Conference of the Rome Statute. Retrieved from https://treaties.un.org/doc/source/docs/RC-Res.6-ENG.pdf

Koh, H. H. (2012). International law in cyberspace. *Harvard International Law Journal Online*, 54, 1.

Kreβ, C. (2006). The crime of genocide under international law. *International Criminal Law Review*, 6(4), 461.

Lewis, J. (2020, February 4). *Election Interference and the Emperor's New Clothes*. Center for Strategic & International Studies. Retrieved from www.csis.org/analysis/election-interference-and-emperors-new-clothes?gclid=EAIaIQobChMI5vfYtZ-36wIVgP3jBx-3YLArAEAAYASAAEgKN-fD_BwE

Mačák, K. (2015). Military objectives 2.0: The case for interpreting computer data as objects under international humanitarian law. *Israel Law Review*, 48(1), 55.

Mačák, K. (2019). On the shelf, but close at hand: The contribution of non-state initiatives to international cyber law. *AJIL Unbound*, 113, 81.

Mačák, K., Gisel, L., & Rodenhäuser, T. (2020, March 27). *Cyber Attacks against Hospitals and the COVID-19 Pandemic: How Strong Are International Law Protections?* Just Security. Retrieved from www.justsecurity.org/69407/cyberattacks-against-hospitals-and-the-covid-19-pandemic-how-strong-are-international-law-protections/

Mansour, H. (2017, January 10). *The 1988 Rohingya Extermination Blueprint*. Human Rights for All. Retrieved from https://merhrom.wordpress.com/2017/01/10/the-1988-rohingya-extermination-blueprint/

McAllister, J. R. (2019–20). Deterring wartime atrocities: Hard lessons from the Yugoslav tribunal. *International Security*, 44(3), 84.

Metzi, J. F. (1997). Rwandan genocide and the international law of radio jamming. *American Journal of International Law*, 91(4), 628–651.

Mozur, P. (2018, October 15). A Genocide Incited on Facebook, with Posts from Myanmar's Military. *The New York Times*. Retrieved from www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

O'Hare, R. (2016, November 1). China Proudly Debuts Its New Stealth Jet It Built 'by Hacking into US Computers and Stealing Plans.' *Daily Mail*. Retrieved from www.dailymail.co.uk/sciencetech/article-3893126/Chinese-J-20-stealth-jet-based-military-plans-stolen-hackers-makespublic-debut.html

Ohlin, J. D. (2009). Attempt, Conspiracy, and Incitement to Commit Genocide. *Cornell Law Faculty Publications*, Paper 24.

Ohlin, J. D. (2020). *Election interference: International law and the future of democracy*. Cambridge University Press.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

OTP. (2017, December 4). Report on Preliminary Examination Activities (2017) – Registered Vessels of Comoros, Greece and Cambodia. *International Criminal Court*. Retrieved from www.icc-cpi.int/Pages/item.aspx?name=2017-otp-rep-PE-Comoros

The Council of Advisers. (2021). Report on the Application of the Rome Statute of the International Criminal Court to Cyberwarfare. Retrieved from www.regierung.li/files/medienarchiv/The-Council-of-Advisers-Report-on-the-Application-of-the-Rome-Statute-of-the-International-Criminal-Court-to-Cyberwarfare.pdf

The Paris Call for Trust and Security in Cyberspace. (2018, November 12). The Paris Call. Retrieved from https://pariscall.international/en/call

Policy Paper on Preliminary Examinations. (2013, November). *International Criminal Court*. Retrieved from www.icc-cpi.int/iccdocs/otp/otp-policy_paper_preliminary_examinations_2013-eng.pdf

Prosecutor v. *Akayesu*, Case No. ICTR-96-4, Judgment (September 2, 1998).

Prosecutor v. *Ahmad Al Faqi Al Mahdi*, Case Information Sheet. ICC-01/12-01/15. Retrieved from www.icc-cpi.int/CaseInformationSheets/al-mahdiEng.pdf

Prosecutor v. *Al Hassan*, Case No. ICC-01/12-01/18-601-Red OA, Judgment on the appeal of Mr. Al Hassan against the decision of Pre-Trial Chamber I entitled '*Décision relative à l'exception d'irrecevabilité pour insuffisance de gravité de l'affaire soulevée par la defense*' (February 19, 2020).

Prosecutor v. *Bemba*, Case No. ICC-01/05-01/08-424, Decision Pursuant to Article 61(7) (a) and (b) of the Rome Statute on the Charges of the Prosecutor Against Jean-Pierre Bemba Gombo (June 15, 2009).

Prosecutor v. *Bemba*, Case No. ICC-01/05-01/08, Judgment (March 21, 2016).

Prosecutor v. *Lubanga*, Case No. ICC-01/04-01/06-A-5, Judgment on the Appeal of Mr. Thomas Lubanga Dyilo against his conviction (December 1, 2014).

Prosecutor v. *Ntaganda*, Case No. ICC-01/04-02/06, Judgment (July 8, 2019). Retrieved from www.icc-cpi.int/CourtRecords/CR2019_03568.PDF

Prosecutor v. *Tadić*, Case No. IT-94-IAR72, Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction (October 2, 1995) ("Tadić case").

Rome Statute of the International Criminal Court. (1998, July 17). UN Doc. A/CONF.183/9*. As amended.

Rona, G. (2003). Interesting times for international humanitarian law: Challenges from the "War on Terror." *The Fletcher Forum on World Affairs*, 27(2), 55.

Roscini, M. (2019). Gravity in the statute of the International Criminal Court and cyber conduct that constitutes, instigates or facilitates international crimes. *Criminal Law Forum*, 30(3), 247.

Rowe, N. C. (2007). War crimes from cyberweapons. *Journal of Information Warfare*, 6(3), 15. Retrieved from https://faculty.nps.edu/ncrowe/iwcrimes.htm

Schense, J., & Carter, L. (Eds). (2017). *Two steps forward, one step back: The deterrent effect of international criminal tribunals*. Torkel Opsahl Academic EPublisher.

Schmitt, M. N. (Ed.). (2017, 2nd edn). *Tallinn manual 2.0 on the international law applicable to cyber operations*. Cambridge University Press.

Schmitt, M. N. (2019, September 16). France's Major Statement on International Law and Cyber: An Assessment. *Just Security*. Retrieved from www.justsecurity.org/66194/frances-major-statement-on-international-law-and-cyber-an-assessment/

Shackelford, S. J. (2017). The law of cyber peace. *Chicago Journal of International Law*, 18(1), 1.

Statute of the Special Court for Sierra Leone. Retrieved from www.rscsl.org/Documents/scsl-statute.pdf. ("Special Court Statute").

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Stubbs, J., & Bing, C. (2019, October 21). Hacking the Hackers: Russian Group Hijacked Iranian Spying Operation, Officials Say. *Reuters*. Retrieved from www.reuters.com/article/us-russia-cyber/hacking-the-hackers-russian-group-hijacked-iranian-spying-operation-officials-say-idUSKBN1X00AK

Trahan, J. (2011). The Rome Statute's amendment on the crime of aggression: Negotiations at the Kampala review conference. *International Criminal Law Review*, *11*(1), 49.

Trahan, J. (2018). From Kampala to New York—The final negotiations to activate the jurisdiction of the International Criminal Court over the crime of aggression. *International Criminal Law Review*, *18*(2), 197.

Trahan, J. (2021). International justice and the International Criminal Court at a critical juncture, in C. Ankersen & W. P. S. Sidhu (Eds.), *The future of global affairs: Managing discontinuity, disruption and destruction*. Palgrave Macmillan.

Trahan, J. (*Forthcoming*). The criminalization of cyberattacks under the International Criminal Court's Rome Statute. *Journal of International Criminal Justice*.

Tsagourias, N. (2012). Cyber attacks, self-defense and the problem of attribution. *Journal of Conflict & Security Law*, *17*(2), 229.

UK Government. (2018, May 23). *Cyber and International Law in the 21st Century*. Retrieved from www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century

UN Charter. (1945, October 24). 1 UNTS XVI.

UN GGE and OEWG. (n.d.). Digital Watch. Retrieved from https://dig.watch/processes/un-gge

U.N. Security Council. (2005, March 31). U.N. Security Council Resolution 1593, U.N. Doc. S/RES/1593.

U.N. Security Council. (2011, February 26). U.N. Security Council Resolution 1970, U.N. Doc. S/RES/1970.

United Nations Treaty Collection. (As of 2020, August 7). Ch. XVIII. Penal Matters. Amendments on the crime of aggression to the Rome Statute of the International Criminal Court. Retrieved from https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XVIII-10-b&chapter=18&lang=en ("States Parties to the Kampala amendment")

U.S. Department of Homeland Security, CISA Cyber & Infrastructure. (Last revised 2019, November 20). *Understanding Denial-of-Service Attacks*. CISA. Retrieved from www.us-cert.gov/ncas/tips/ST04-015

Warrell, H., Seddon, M., & Manson, K. (2020, February 20). Russia Military Unit Accused of Georgia Cyber Attacks. *Financial Times*. Retrieved from www.ft.com/content/14377b84-53e3-11ea-90ad-25e377c0ee1f

Whiting, A. (2015, July 20). *The ICC Prosecutor Should Reject Judges' Decision in Mavi Marmara*. Just Security. Retrieved from www.justsecurity.org/24778/icc-prosecutor-reject-judges-decision-mavi-marmara/

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

8

Trust but Verify: Diverse Verifiers Are a Prerequisite to Cyber Peace

Rob Knake and Adam Shostack[*]

"Trust but verify." Students of history and readers of a certain age will recall those words being spoken by former US President Ronald Reagan. His argument was that peace required verification mechanisms so that each side could be confident in the actions of the other side. There are important lessons for cyber peace. While Reagan was speaking in the context of strategic nuclear arms control, many papers have been devoted to the difficulties of arms control in the cyber domain (Maybaum and Tölle, 2016). Cyber weapons do not require the large physical infrastructure of nuclear programs and can far too easily be kept secret to allow for meaningful validation of adherence to arms control commitments. Moreover, many "cyber weapons" are dual use in nature, being deployed for the administration of computers and networks, or for security testing. Yet, as we discuss in this chapter, arms control is only one area in which verification is an important tool for maintaining international peace in cyberspace and other domains.

This chapter starts with a discussion of the role played by verifiers in peace. We discuss some of the many types of verifiers, and how those whose roles are outside the formal political process can help to construct peace. Many of these have scientific or investigatory roles whose work informs the state of the world. There are interesting models in aviation, including not only the National Transportation Safety Board (NTSB) but also a variety of others including institutions dedicated to telemetry analysis and near miss analysis. We examine each and suggest how a cyber equivalent could contribute to our understanding of the state of the world and in doing so, support peace.

1 THE NEED FOR VERIFIERS IN CYBERSPACE

A state of peace is a social construction. Peace exists because all parties act as if it exists, but it can be broken or threatened by the actions of any party. As long as peace and a belief in peace exists, it acts as an inhibitor to the initiation of violence,

[*] The authors would like to thank Scott Shackelford, Steve Luczynski, and the participants in the colloquia for providing helpful comments and feedback on the draft of this chapter and Ben Lefkowitz for his research and editorial assistance.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

because peace is worth preserving. Parties inhibit their activities to maintain peace (or they act to break it). We take these ideas as axiomatic to allow us to investigate the idea of diverse verifiers and investigate several categories of verifiers whose existence would support the construction of peace. Both the construction and effects of peace have many aspects that are explored elsewhere in this volume, particularly in Chapters 1–3.

War and peace are frequently paired with terms of probability, duration, and time: An impending war, an uneasy truce, a stable peace. If people are uncertain about the existence of peace, if they are uneasy about it, then their willingness to make threats, to plan to carry out threats, and to impose their will on others will be higher. If societies are worried about a lack of peace they will invest in security. Building walls and forts takes substantial resources and takes those resources away from other possible investments. They will invest in arming, training, and maintaining military forces. In times of peace, those investments are reduced. The frames assigned to such things (the "peace dividend," "to maintain peace, prepare for war," and the like) are usually normative and closely relate to the speaker's belief in the stability and longevity of peace. A more widely shared belief that the world is at peace and that peace is stable will free resources for nondefense spending. To the extent that such a belief is accurate and well founded, those other areas of spending will reflect the desires (rather than the fears) of the public. Wide sharing of a belief in peace will be enhanced if many groups with different perceived motives are reporting similar things. Contrariwise, if some sources are reporting signs of war and others are not, there will be disagreement over spending.

A cyber peace dividend might consist of several components, including reduced corporate investment, reduced national investment, and reduced cost to the general public. Today, a widely cited rule of thumb is that commercial entities spend about 8 percent of IT budgets on security (Nash, 2019). Not all of that could be reclaimed by a cyber peace dividend. National investments by governments include both attack and defense. The former is easier to unilaterally reduce. We note, in passing, that the offense budgets are often "black budgets" and hard for outsiders to understand. The cost to the public is a mix of anxiety and the inhibition of productive work because security is hard.

In 2020, there is extreme distrust both across and between societies. The Trump administration announced that the United States would withdraw from the World Health Organization (WHO), a move that the Biden administration reversed. The United Kingdom has withdrawn from the European Union. Many people are refusing to wear masks, refusing to believe in climate change, the list goes on. Attacks on the credibility of news organizations ("it's fake news") augment and bolster other attacks on credibility. In order to overcome this distrust, the world would be better with a series of neutral, trustworthy, and trusted institutions that are less subject to political or market forces and must adhere to strict protocols for verifying the claims of actors in cyberspace. "Governments and diplomats," as Roger Hurwitz

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

(2012) notes, "… have been less clear in recognizing how foundational public trust is for cyberspace." Similarly, Elinor Ostrom has commented that "trust is the most important resource" (ESCOTET Foundation, 2010). In that spirit, diverse verifiers are the soil in which trust grows.

We look for inspiration to aviation. Among the reasons to look to aviation is that while aviation is inherently risky, deeply technical, and still relatively new, it has evolved into a set of trusted and trustworthy institutions. In addition, other research projects we have done over the last few years have familiarized us with the institutions there, and on consideration they seem to be perhaps both interesting and inspirational.

2 BUILDING OFF OF THE AVIATION MODEL

In other contexts, international mechanisms exist to investigate claims of activity that violate international agreements or norms of behavior. Interpol and the International Criminal Court both investigate allegations of war crimes and human rights violations. The International Atomic Energy Agency (IAEA) investigates violations of the nonproliferation treaty. Given the limitations we note above on applying the arms control model to cyberspace, a better analogy than nuclear site inspections may be international civil aviation. In the domestic context, the authors have separately and collectively promoted the development of cyber incident investigations, modeled on the National Transportation Safety Board's process for investigating aviation incidents and the processes for sharing "near misses" within the aviation community. In concert with the development of national mechanisms for investigating cyber incidents, the international community is also in need of international mechanisms to coordinate and referee international cyber incidents involving multiple states.

For international aviation incidents, the Convention on International Civil Aviation (1994) dictates that the jurisdiction of the crash site will have primary responsibility but allows that jurisdiction to cede authority to a different authority. Such arrangements are managed through the International Civil Aviation Organization (ICAO), the organization established by the convention. In the case of the Malaysian Airlines Flight 17, which was shot down by Russian-backed rebels over Ukraine on July 17, 2014, Ukraine delegated the Netherlands to conduct the investigation given that the flight originated in Amsterdam and had a large number of Dutch citizens onboard (Parker and Olearchyk, 2014). The decision may also have created the perception of improved capability and objectivity by bringing in a third country that was not embroiled in the ongoing conflict to conduct the investigation. In the case of Malaysian Airlines Flight 370, which disappeared over the Indian Ocean on March 8, 2014, Malaysia assembled a Joint Investigative Team of experts from Malaysia, China, the United Kingdom, and the United States, led by an independent investigator under ICAO standards.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

In contrast, when international cyber incidents occur, investigations are conducted in an ad hoc manner, usually under the authority of the victim state or by private firms. The findings of such investigations are often the subject of political machinations by the victim company or organization who may wish to avoid negative market reactions for failing to prevent the incident; by the victim's government, which may either seek to downplay or promote the narrative depending on the geopolitical concerns of the moment; and, of course, by the attacker or the attacker's country. In the vast majority of cases, however, no investigative report is ever published. Incident response will be carried out for the purposes of containing an ongoing incident, recovering systems, and preventing future incidents at the victim company. Incident handlers are not, however, in the business of fact finding and reporting so that lessons can be learned and, thus, similar incidents being prevented at other companies.

Some incident handlers generate or contribute to a product labelled "threat intelligence." These "feeds" are often commercial and include the attacker's given names like "Dynamite Panda" (MITRE ATT&CK, 2020). Many times, these products include attribution information, such as "this group uses these tactics," or "the Panda set of attackers are Chinese Government affiliated." The quality of these products have not fared well under scrutiny (Bouwman et al., 2020).

On attributing an attack to a specific state, attribution is also typically carried out in ad hoc manner, as was discussed more fully in Chapter 7. Cybersecurity firms may choose to attribute the incidents they discover, or prevent the actions of specific states, if they see it in their commercial interest, or believe that they have a patriotic duty to do so. More often than not, however, cybersecurity firms will choose to avoid attributing activity to a specific nation state so as not to hurt their commercial prospects in that state, or to avoid becoming a target themselves of that state. When national governments make a claim attributing malicious cyber activity to an adversary state, those claims are typically rebuffed by the accused state and largely ignored by the international community.

3 BACKGROUND: HISTORICAL INCIDENT INVESTIGATIONS

In the United States, investigations of cyber intrusions are typically conducted by private, for-profit cybersecurity firms. In rare cases, when a significant incident occurs, the federal government will investigate and report out on the incident. When the incident involves a federal computing system, such as the incident at the Office of Personnel Management (OPM) in 2015, Congress may investigate. In other cases, Congress asks the Government Accountability Office (GAO) to investigate. These reports are often slow to be produced and can be highly political in nature. While they may provide lessons learned to the cybersecurity community, that is not their primary purpose. Instead, their goal is to assign blame, sometimes in a highly partisan fashion. In the case of the OPM data breach in 2014, the House Oversight and Government Report Committee issued a 241-page report on the incident titled "The OPM Data Breach: How the Government Jeopardized Our National Security for

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

More than a Generation." While the report provides a comprehensive review of the incident that is valuable from a historical context, its partisan tone undermines its legitimacy as an even-handed fact-finding effort. Its timing, two years after the incident and a month before a hotly contested presidential election, also led to questions about its motivation and purpose.

On the international front, as with the downing of Malaysian Airlines Flight 17 in the air domain, Ukraine has proven to be the focus of significant international conflict within the cyber domain due to the protracted conflict between Russian-backed separatists and the western Ukraine government. Offensive cyber operations that were conducted against electric sector targets caused widespread power outages on two occasions. Ukraine was also the target of the NotPetya malware attack. Given the global spread of NotPetya and international concern over attacks on critical infrastructure, this analysis will focus on the attacks on the power grid. In the first of those incidents (in December of 2015) offensive cyber operators took thirty substations and two power distribution centers offline. The Ukrainian government sought international assistance to investigate the matter. According to reporting by Wired Magazine (Greenberg, 2017), the investigation into the incident was conducted by Ukrainian officials with the assistance of the US Federal Bureau of Investigations and the US Department of Homeland Security. At least two private sector experts were brought in to assist the investigation. They were Robert Lee, a former National Security Agency technical operator and CEO of the industrial control systems security firm Dragos, and Michael Assante, the former chief information security officer (CISO) for the North America Electric Reliability Corporation. Both Lee and Assante were also instructors at the private SANS Institute.

Following the investigation, Lee and Assante published a publicly available report, "Analysis of the Cyber Attack on the Ukrainian Power Grid" (2016), under the auspices of the SANS Institute and the Electricity Information Sharing and Analysis Center (E-ISAC). That report addressed one of the two main purposes for conducting such an investigation, relating to other security professionals what happened so that lessons could be learned to prevent other, similar incidents in the future. It did not, however, address attribution of the attack. The Ukraine government asserted that the attack was carried out by Russia, but no international body validated that claim and the Ukrainian government offered no proof to substantiate the claim. For its part, the US government has never publicly attributed the attack to Russia, but leaks to the media have substantiated the claim (Park et al., 2017).

While the 2015 attack could have been the launching point of an effort to investigate incidents at critical infrastructure and disseminate lessons learned, no such virtuous cycle of process development and ongoing improvement began. When the Ukrainian power grid was attacked a second time, in December of 2016, the incident garnered far less attention. A standout example of dissemination of findings following a cyber incident was the March 2019 breach of Norsk Hydro, a Norwegian aluminum maker. Norsk Hydro made the unprecedented decision to be fully transparent

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

about the incident, hosting web conferences to disseminate findings to the security community. In this incident, Microsoft's Detection and Response Team led the response and authored the main report on it (Briggs, 2019).[1]

4 INVESTIGATING DOMESTIC INCIDENTS: THE NEED FOR A NATIONAL CYBERSECURITY BOARD

When a major security incident happens, victims are strangely incented to lavish praise on the attackers. After all, there is little shame in being hacked by the pros – "how were we supposed to fight the Russians?" So, some attacks that were performed by criminals or even teenage hackers will be blamed on professionals. If the Acme Company blames the KGB, who is to contradict them? From where do we get our facts? This misattribution is not harmless. The act of blaming the Russians (the Israelis, the Chinese, and the North Koreans) undercuts our assurance of a state of peace.

An investigatory board could help provide those facts. Reports from the NTSB, for example, are seen as authoritative and trustworthy. An investigatory board that invested in gaining and maintaining a reputation for competence could be a substantial counterbalance to organizations spreading self-serving claims. For example, a cyber board could conduct an investigation and release a report that assessed the sophistication displayed by an attacker on a scale from "not sophisticated" to "exceptionally sophisticated." It could assess the idea that an attack was carried out by a nation state or the reliability of a claim that it was a particular nation state.

As this is being drafted, the United States, United Kingdom, and Canada released a joint statement claiming that Russian Intelligence is trying to steal vaccine information (NCSC et al., 2020), but such statements are unusual. The process for releasing intelligence information is opaque. Is the absence of such an announcement the result of peace or a geopolitical decision by intelligence agencies to withhold information?[2] By credibly communicating facts, a cyber board could be a stabilizing force for peace.

4.1 *Why Do We Not Already Have a Cyber NTSB?*

This subsection starts with a brief summary of what the NTSB does, examines some of the objections to a cyber analog, continues with some of the ways those objections might be addressed, and ends with some practical, achievable steps to create a cyber version NTSB. The NTSB is best known for investigating *accidents* in *aviation*.

[1] We do not mean to cast aspersions on Microsoft, but having the creator of the operating system that was attacked may introduce bias.

[2] An intelligence agency might withhold information to protect sources and methods, or to continue an operation to meet additional objectives.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Aviation is a regulated sector. For an airplane to exist (in the United States) requires permission from the FAA; taking off requires a qualified pilot at the controls before leaving an airfield. Each of these is a term of both law and art and, while exceptions exist, these many constraints also act as constraints on the NTSB. An accident is something that leads to the death or injury of someone on a plane, or meaningful damage to one, and these are usually prerequisites to, and provide scope for, an investigation.

The first call we know of for a cyber investigations board was in the 1991 National Research Council report, *Computers at Risk*. Yet no such board exists thirty years later, and the reason, we think, is primarily industry opposition.[3] The core of that opposition is concern. No one wants to have their actions judged with 20/20 hindsight. No one wants to have their innovation judged by those who've never operated a business or been responsible for a profit and loss account. And while such judgements may or may not be real, the perceived threat inhibits the creation of such a board. In contrast, the NTSB was created when accidents in aviation were frequent, and those accidents inhibited the growth of the sector. The aviation industry came together in support of an investigatory body. In contrast, the technology sector seems to be generally opposed. It may be that there is also support, for example, from the insurance industry, but such support has not caused a cyber version of the NTSB to come into existence.

The fear of being judged can be a real problem. An interesting quote from *Roving Mars* (Squyres, 2005 discusses the choice to launch the Mars Exploration Rovers (Spirit and Opportunity)). Before we reach this scene, there was one prelaunch review board after another, examining the engineering choices that had been made:

> Chris Scolese, Ed's deputy, was still in the room, and he explained what had happened. Chris is an engineer, and he has managed space flight projects. What Chris knew is that practically every spacecraft that's ever flown has had some kind of weird problem that popped up once or twice during testing, never to be seen again. You have to take some risks in this business, and the risk we were taking with the transponder was lower in Chris's judgement than the risks we'd already decided we were willing to take on launch day and landing day. Chris had told Ed that he thought we should fly, and Ed had accepted Chris's advice. But it had been a tough call by both of them.

With 20/20 hindsight, Scolese's decision was right, but imagine if the rocket had blown up. Was "you have to take some risks" and "the risks were lower with the transponder" really justifiable? The prospect of such questioning inhibits experimentation and risk-taking. Sometimes that inhibition is appropriate. We would all agree that it is important to have test systems that mirror the production system as

[3] There have been many analogies made to such a system, under a variety of acronyms. For this chapter, we generally will refer to such things as a board, an investigations board, or even a cyber investigations board, using the terms interchangeably with specifics to improve readability.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

closely as possible, and to test with those systems, right? Take a moment to think and see if you agree. Sometimes that inhibition is appropriate. That being said, progress requires innovation and experimentation, and blame and second-guessing inhibit such experimentation.

As it turns out, the real world is a strange and complex place. It turns out that companies like Facebook and Netflix have moved to a practice of rolling out changes slowly across subsets of their production systems. This practice is often derisively called "testing in production," which was a shocking strategy when these companies first admitted to it (Mappic, 2011). If those trying it had been worried about an external review board, they might have been prevented from experimenting. Testing in production is now accepted practice; it is considered by some to be a leading approach.

Industry concerns about having their practices judged are strong and real, as is the regular reinvention of the idea. It may be that there are ways to square this circle.

4.2 *Getting to a Cyber NTSB*

To stand up to a cyber incidents investigation board, we must balance the real and perceived concerns with an understanding of the myriad benefits, which include the ability to learn from the misfortune of others and to support the construction of peace. A board does not have to investigate everything to be useful to the cause of peace. The NTSB's role is strictly constrained to accidents involving transportation; thus, a cyber version could be created in a way that aids in peace while addressing corporate concerns.

For example, such a board could initially limit its investigations to breaches involving US Government computers and limit its investigation of more complex incidents to the government computer subset of those cross-entity incidents.[4] As the capability of the organization grows, and as processes mature, the scope could be expanded to other critical infrastructures or other organizations could be created for this purpose. Today, these might be investigated by the FBI, and the attackers might be the subject of surveillance or other operations by intelligence agencies. Each of these agencies has limited resources, and different goals. Managing the overlap of such investigations may carry some complexity. However, this is a reality of complex incidents. For example, the Air Force already imposes such complexity on itself. Accidents are investigated by both a Safety Investigation Board and an Accident Investigation Board, each with different goals (Air Combat Command, 2013).

Another key question area would be the ability of a board to compel participation by either or both an organization and specific staff. Obviously, the participation of the victim organization is important, but to what extent is it expected and

[4] One of our reviewers commented that limiting to "just" US government computers seems quite narrow. We agree, and it would be much broader than what we have today.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

reasonable? What about their staff? To what extent should an investigations board be able to compel participation from suppliers to that victim? Would Microsoft, Google, and Amazon need staff dedicated to answering the board when their products are involved in a breach? Would investigators be limited to "what's in the manual" or can they delve into product design decisions?[5] Even with regard to the manual, it is not always obvious what section of a complex product's technical documentation is relevant. The two volumes of the latest edition of "Windows Internals" (Yosifovich et al., 2017) comprise 1,568 pages, and those are books. The more voluminous technical documentation is now largely online and updated frequently. What is a reasonable expectation of an operator of such systems? These questions are not insurmountable, but some versions of them need to be addressed to move proposals forward.

What about the participation of staff? Can that be compelled? What about the right against self-incrimination? As we write this, Uber's former Chief Security Officer has just been charged with obstruction of justice. What are the expectations for staff of a breached organization in terms of participation in an investigation? Is it "answer three questions by email" or "be deposed for a day or more?" How are software development staff to be trained, and whose staff would receive training? For example, the Air Force delivers annual training to pilots on the various investigations that will happen after an accident.

4.3 *What Could a Cyber NTSB Do for Peace?*

Calls for a cyber investigations board have traditionally focused on learning and disseminating lessons from incidents. This is inherently useful in the creation and preservation of cyber peace because it makes future attacks more difficult. And there are many other ways in which a board could support the cause of peace, including the following:

- Publishing lessons learned reports (as opposed to sharing them under NDAs)
- Bring different goals to incident investigation
- Investigating more/different cases than police or intelligence agencies
- Provide attribution with different biases
- Report on the state of the world
- Provide international assistance
- Support a construction of peace

The primary reason for previous calls for a cyber investigations board has been to find and distribute lessons. The incredible safety record of aviation is commonly attributed to these and other learning systems. An investigations board could

[5] Even suggesting this discomforts the author, Shostack. Having each of the product tradeoffs judged raises issues discussed elsewhere.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

establish consistency and credibility, and stand in complement to the information released by police and prosecutors. That information is focused on literally "making the case" for prosecution and conviction, rather than learning lessons or informing. Analysis that is designed to be objective could better support peace by informing debate about the state of the world. It could potentially do so in a larger set of cases if the investigators are not required to testify, be subjected to cross-examination, and perform other tasks in the judicial system. The cases that a board investigates might be quite different than the ones that the police investigate. (There would need to be a deconfliction/equites process to ensure that investigations did not accidentally cross paths with other investigations. That process, like all the others, requires training for the involved participants.)

A board could provide attribution information about cases with a different authority than either private or prosecutorial analysis. Such analysis might be read with less skepticism or read with different skepticism by different parties, providing information that either supports or undercuts the construction of peace through a better understanding of the state of the world.

In addition to information about specific attacks, additional high-quality information about the frequency and intensity of international attacks would illustrate the state of the world at a given time and add information about the actors who are violating the peace, increasing the likelihood that they would be either caught[6] or meaningfully made to take the blame for their actions.

The NTSB provides help and assistance to air crash investigations around the world. It would not be unreasonable to expect that once a board had established itself and its competence, it could, when asked, help investigate "important incidents" outside of the federal government, including state and local governments, as well as, perhaps, private enterprises. This assistance to entities within national borders could raise the cost of attacks via exposure. International assistance could be an act of goodwill, bolstering peace.

Additionally, a stream of analytic reports that establish norms and expectations would inform industry's position on the impact of investigations. While it is reasonable to think that more data would aid in the understanding of the state of the world as was described in Chapter 3, it is similarly reasonable to think that most industry benefits from peace and trade.

5 A SYSTEM FOR REPORTING NEAR MISSES

The NTSB is the best known of a polycentric constellation of aviation safety programs which complement and overlap to make hurtling through the air at hundreds of miles per hour incredibly safe. There are others including the Aviation

[6] Methodological analysis of incidents might cause attacks that had been attributed to criminals to be correctly attributed to state actors, or vice versa.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Safety Reporting System (ASRS) and the Aviation Safety Information Analysis and Sharing System (ASIAS). One of the authors (Shostack) has argued at length for a Cyber Security Reporting System (CSRS),[7] and we believe that such a system could also enhance and preserve peace (Bair et al., 2017). Before discussing near misses at some length, we will first briefly explain the ASIAS system, and some of the limits an ASIAS analog would face. This helps illustrate the value of an ASRS-like system.

5.1 ASIAS: *Telemetry Analysis*

The ASIAS program collects telemetry from aircraft in operations, analyzes it, and reports back to the operators. For example, if flights operated by one airline have substantially different wing flutter than those operated by other airlines from that same airfield, then that might be interesting for each airline to know. Our ability to compare telemetry is built on a scaffolding of similarities. Aircraft and their components are made by a small number of manufacturers. The operational systems are defined by flights of a limited number of types (general, cargo, and military) from one field to another. This leads to similarity between the telemetry each emits. Computer systems run a far more varied set of workloads. A mail server might run on Windows, Linux (Ubuntu, Debian, RedHat, etc.), FreeBSD, OpenBSD, or others (McKusick et al., 1996). The mail software might be sendmail, postfix, qmail, or Exchange, or even Gmail or Hotmail, which are (reputedly) unique software. Each of these operating systems and mail packages logs differently. Similarly, there is diversity in each "stack" of software, and that software delivers diverse values.

Despite this diversity, aggregated analysis of attacks could produce useful information. For example, if logs of rejected emails were collected, then we could learn about spam campaigns. There is a difference between mail from northeastem.com going to northeastern.edu and it going to shostack.org. On first blush, the former is much more likely to be a targeted campaign, and the latter to indicate a broad spamming campaign. But if we gathered rejection data from many recipients about email domains, we could tell recipients about the unusual campaigns they receive. Unusual might be determined algorithmically based on those whose sending domains are unusual, and there are standard computer science techniques that would help determine what counts as unusual relative to each recipient.[8] The data sent back to participants could motivate their participation, and the agency performing the analysis could provide information about the state of conflict in the world and possibly between states and semi-state and nonstate actors.

[7] Since there are fewer calls for such a thing, we will use the CSRS acronym.

[8] There are standard techniques that could be applied, for instance, term frequency/inverse document frequency, or "small edit distance."

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

5.2 *ASRS: Near Miss Reporting*

We believe we can develop broader, and perhaps less expected lessons, from a cyber version of ASRS. In aviation, if there is an incident, then anyone involved can submit a short, two-page form to the ASRS, operated by NASA.[9] An incident is anything short of an accident, which, again, is the death or injury of a person or damage to an aircraft. The reports go to NASA to isolate them from accidental disclosure to the regulators. (There are important additional protections in both law and agreements between NASA and the FAA.) NASA ingests the reports, analyzes them, and publishes data that are carefully anonymized.[10] NASA also sends back a receipt. The reporter can use that receipt to demonstrate "evidence of constructive engagement" in a disciplinary proceeding. This evidence is one of the factors that the FAA takes into account in its administrative law proceedings. This incentive, which might seem small, adds to each participant's desire for a safe aviation system and is enough to motivate roughly 100,000 reports each year to the ASRS (ASRS, 2019).

5.3 *Cyber Near Misses and What We Might Learn*

Near miss reporting, both within and between organizations, is an important building block in safety programs in a great many industries. Similarly, many of these programs use blamelessness as a tool to demonstrate their prioritization of learning over retribution.

The nature of near misses in cybersecurity makes them easier to report and discuss, and that eases open doorways to understanding the state of the world. The sorts of things we might understand include (but are not limited to) attacks that progress too close to a meaningful target or attacks that gain the interest of investigators for their distinctiveness. In doing so, near-miss reporting makes more measurable what is commonplace and effective, such as phishing and the techniques in use. These are nominally reported on, but what's almost working can be lost in the noise.

We can learn useful things about what works to protect, detect, and respond to problems by tracking which tools are reliably reported for each. Such analysis can be broad and helps us to better preserve peace by prioritizing effective defenses. For example, while the NIST CSF contains over 900 controls (Reciprocity Labs, 2019),[11] the Australian Signals Directorate recommended a "top 4," now transformed into an "essential eight" (Coyne, 2017).[12] Even if we believe that the controls in each set

[9] The form can be found at https://asrs.arc.nasa.gov/report/electronic.html

[10] The anonymization has both a technical component and a review component.

[11] The NIST CSF is the National Institute for Standards and Technology's Cyber Security Framework, one of the primary ways the United States specifies the cybersecurity defenses (controls) that organizations are expected to deploy and maintain.

[12] If the Australians double their list every three years, it will still take till roughly 2042 before they're closing in on 900 controls.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

are at different levels of abstraction, and thus each of the eight represents a dozen in the NIST set, there remains a massive difference in the control recommendations. Either one of these standards is missing crucial controls, or the other standard includes investments that do not do very much good.[13] Knowing what does not work can be an important step forward. Stopping ineffective investments makes room for new ones. So, both positive and negative reports can be useful. A mix allows for interesting science: Why does measure A work for some organizations but not others?

5.4 *The Contribution of a CSRS to Cyber Peace*

The first contribution of a CSRS to peace would be the ability to improve defenses or to reduce costs without reducing the quality of defenses. The former makes attacks harder, and the latter allows us to invest in other things. Today in cyber warfare, the attacker has tremendous advantages. Improving the effectiveness of defenses would shift the balance somewhat. Making attacks more difficult, more likely to be detected, or more attributable would shift the logic against launching attacks and thus contribute to peace.

The second contribution could be an assessment of attacker activity. If a CSRS-adjacent body had access to confidential descriptions of "tactics, techniques, and procedures," then it could analyze near miss information to report on rates of attacks or attack intensity.[14] This would be a very different function than aviation's ASRS, but streams of near miss information in cybersecurity could be leveraged for this. Such variation may cause problems for multinational companies reporting to local authorities.

6 AN INTERNATIONAL MECHANISM TO INVESTIGATE AND ATTRIBUTE CYBER INCIDENTS

Building off of the ICAO model, what is needed in the international context is a mechanism for requesting international support for investigating significant cyber incidents. These investigations would be carried out for a dual purpose. First, they would provide a standard process and rapid timeline for disseminating findings useful to cyber defenders. Second, they would provide a means for determining attribution and releasing such findings to the public, allowing other international bodies

[13] There is another possibility, which is that they are aiming at different levels of security, but since we have no measure of what that means, we exclude it.

[14] TTPs and "indicators of compromise" are things such as domains used by attackers, email subjects, IP addresses, and malware identifiers. They are useful for detecting and grouping attacker behavior. They are often kept close to the vest to prevent attackers from becoming aware that defenders are using them. Collective reporting of an analysis might be easier to report on than specific comments like "the Acme corp managed an attack by the Drunken Bear APT group."

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

to censure or penalize the offending state. These findings could also serve as the basis for organizing coalitions of governments to sanction or otherwise condemn the actions of the offending state should international institutions fail to act.

At this stage, rather than funding a standalone organization to investigate international cyber incidents, a more modest approach would be to establish a concept of operations for how such investigations should take place and who should take part in them. As in the successful example of the 2015 Ukraine investigation, such investigations will need to rely heavily on private sector expertise. Particularly in the area of industrial control systems, expertise on the security and forensic methods for such systems is exceedingly rare. Thus, keeping experts with the knowledge to carry out these investigations on the sidelines while waiting for the phone to ring would not be practical. Instead, ad hoc teams should be formed at the behest of the victim state. These teams would be invited to investigate and issue initial findings in a rapid fashion, followed by a comprehensive final report issued by the international body sponsoring the effort.

Some of these functions might be picked up by a "Cyber Peace Corps," as discussed elsewhere in this volume, including in the essays section. But such a group, with room for everyone, carries a different function and requires a different culture from an organization with strong leadership focus on producing investigative reports. A Peace Corp could be a feeder to such an investigative body, helping to respond to problems, preserving evidence, and bringing forward interesting cases.

On determining attribution, significant conclusions can typically be achieved by comparing the tradecraft of the attacker to other known historic incidents. This process has led ESET (2016) and Dragos (2017), among others, to conclude that the team behind the Ukraine attacks was the same team behind the attacks on the Democratic National Convention and other political targets in the lead up to the 2016 US presidential election. Thus, without the benefit of national intelligence capabilities, investigators should be able to make preliminary conclusions on attribution. Intelligence agencies could then provide their own findings to the team, agreeing to release some, all, or none of the evidence uncovered through intelligence collection to the public. This process would allow for sources and methods to largely be protected, while providing an independent verification mechanism of the claims.

CONCLUSION

In this chapter, we have argued that trusted verifiers are essential for cyber peace. By creating trusted national mechanisms for investigating cyber incidents, lessons learned can be shared with the wider community and confidence that problems that caused one incident can be corrected elsewhere before more such incidents occur. By creating trusted verifiers for near misses, all members of the cybersecurity community can provide telemetry to determine the current level of

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

hostility in cyberspace. With a strong international mechanism for investigating significant cross border cybercrime, determining lessons learned, and attributing malicious activity, more consequences can be created for states that engage in such activity. As norms of conduct in cyberspace are developed, it is essential that verifiers are enabled at multiple levels to ensure that they are being upheld, and when they are not to verify that claims of malfeasance are proved true and taken seriously. Trust but verify is, now more than ever, essential to the preservation of peace.[15]

BIBLIOGRAPHY

Air Combat Command. (2013, January 11). *Air Force Safety* and *Accident Board Investigations*. www.acc.af.mil/About-Us/Fact-Sheets/Display/Article/199117/air-force-safety-and-accident-board-investigations/

Aviation Safety Reporting System. (2019, July). *ASRS Program* Briefing. National Aeronautics and Space Administration. https://asrs.arc.nasa.gov/docs/ASRS_ProgramBriefing.pdf

Bair, J., Bellovin, S. M., Manley, A., Reid, B., & Shostack, A. (2017). That was close: Reward reporting of cybersecurity near misses. *Colo. Tech. LJ*, *16*, 327.

Bouwman, X., Griffioen, H., Egbers, J., Doerr, C., Klievink, B., & van Eeten, M. (2020). A different cup of {TI}? The added value of commercial threat intelligence. In 29th {USENIX} Security Symposium ({USENIX} Security 20) (pp. 433–450).

Briggs, B. (2019, December 16). *Hackers hit Norsk Hydro with ransomware. The company responded with transparency*. Microsoft. https://news.microsoft.com/transform/hackers-hit-norsk-hydro-ransomware-company-responded-transparency/

Cherepanov, A., & Lipovsky, R. (2017, June 12). *Industroyer: Biggest threat to industrial control systems since Stuxnet*. WeLiveSecurity. www.welivesecurity.com/2017/06/12/industroyer-biggest-threat-industrial-control-systems-since-stuxnet/

Committee on Oversight and Government Reform. (2016, September 7). *The OPM Data Breach: How the Government Jeopardized Our National Security for More than a Generation*. U.S. House of Representatives. https://republicans-oversight.house.gov/wp-content/uploads/2016/09/The-OPM-Data-Breach-How-the-Government-Jeopardized-Our-National-Security-for-More-than-a-Generation.pdf

Convention on International Civil Aviation. (1994, December 7). International Civil Aviation Organization (ICAO), U.N. Document 7300. www.icao.int/publications/pages/doc7300.aspx

Coyne, A. (2017, February 6). *Overhaul of ASD's Top 4 cyber threat strategies*. itnews. www.itnews.com.au/news/drastic-overhaul-of-asds-top-4-cyber-threat-stategies-449787

Dragos. (2017). *CRASHOVERRIDE: Analysis of the Threat to Electric Grid Operations*. www.dragos.com/wp-content/uploads/CrashOverride-01.pdf

E-ISAC. (2016). Analysis of the cyber attack on the Ukrainian power grid. https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf

ESCOTET Foundation. (Fall of 2010, reproduced fall of 2020). *Interview with Nobel Laureate Elinor Ostrom*. https://escotet.org/2010/11/interview-with-nobel-laureate-elinor-ostrom/

[15] Since this chapter was written, the authors have released a technically focused report on the subject of learning systems: Robert Knake Adam Shostack Tarah Wheeler, Learning from Cyber Incidents: Adapting Aviation Safety Models to Cybersecurity, Belfer Center for Science and International Affairs, Harvard Kennedy School, November 12, 2021, www.belfercenter.org/learning-cyber-incidents

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Greenberg, A. (2017, June 20). How an entire nation became Russia's test lab for cyberwar. *Wired*. www.wired.com/story/russian-hackers-attack-ukraine/

Hurwitz, R. (2012). Depleted trust in the cyber commons. *Strategic Studies Quarterly*, 6(3), 20–45. www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-06_Issue-3/Fall12.pdf

Lemos, R. (2020, August 14). Research casts doubt on value of threat intel feeds. *Dark Reading*.www.darkreading.com/threat-intelligence/research-casts-doubt-on-value-of-threat-intelfeeds/d/d-id/1338676

Mappic, S. (2011, December 6). Why testing in production isn't as stupid as it sounds. Appdynamics blog, www.appdynamics.com/blog/product/why-testing-in-production-isnt-as-stupid-as-it-sounds/

Maybaum, M., & Tölle, J. (2016, May). Arms control in cyberspace-architecture for a trust-based implementation framework based on conventional arms control methods. In 2016 8th International Conference on Cyber Conflict (CyCon) (pp. 159–173). IEEE.

McKusick, M. K., Bostic, K., Karels, M. J., & Quarterman, J. S. (1996). *The design and implementation of the 4.4 BSD operating system (Vol. 2)*. Addison-Wesley.

MITRE ATT&CK. (2020, March 30). *APT18*. https://attack.mitre.org/groups/G0026/

Nash, K. (2019, December 30). Tech chiefs plan to boost cybersecurity spending. *The Wall Street Journal*. Retrieved from www.wsj.com/articles/tech-chiefs-plan-to-boost-cybersecurity-spending-11577701802

National Cyber Security Centre (U.K.), Communications Security Establishment (Canada), & National Security Agency (U.S.A.). (2020). Advisory: APT29 targets COVID-19 vaccine development. United States Department of Defense. https://media.defense.gov/2020/Jul/16/2002457639/-1/-1/0/NCSC_APT29_ADVISORY-QUAD-OFFICIAL-20200709-1810.PDF

National Research Council. (1991). *Computers at risk: Safe computing in the information age*. National Academy Press.

Park, D., Summers, J., & Walstrom, M. (2017, October 11). *Cyberattack on Critical Infrastructure: Russia and the Ukrainian Power Grid Attacks*. The Henry M. Jackson School of International Studies: University of Washington. https://jsis.washington.edu/news/cyberattack-critical-infrastructure-russia-ukrainian-power-grid-attacks/

Parker, A., & Olearchyk, R. (2014, July 21). Netherlands to lead MH17 investigation. *The Financial Times*. Retrieved from www.ft.com/content/19c29f34-10e1-11e4-b116-00144feabdc0

Reciprocity Labs. (2019, December 10). *What Are NIST Controls and How Many Are There?* https://reciprocitylabs.com/resources/what-are-nist-controls-and-how-many-are-there/

Squyres, S. (2005). *Roving mars: Spirit, opportunity, and the exploration of the red planet*. Hachette Books.

Treaty on the Non-Proliferation of Nuclear Weapons, 1970.

Yosifovich, P., Solomon, D. A., & Ionescu, A. (2017). *Windows internals, part 1: System architecture, processes, threads, memory management, and more*. Microsoft Press.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

9

Building Cyber Peace While Preparing for Cyber War

Frédérick Douzet, Aude Géry, and François Delerue

Since President Macron's launch of the Paris Call for Trust and Security in Cyberspace in the fall of 2018,[1] amidst the collapse of international cyber norm discussions in June 2017, the international community has contemplated and launched multiple initiatives to restore a multilateral dialogue on the regulation of cyberspace in the context of international security. In December 2018, two resolutions were adopted by the United Nations General Assembly to set up two processes on progress in information and telecommunications in the context of international security: The sixth Group of Governmental Experts (GGE)[2] on the subject and a new Open-Ended Working Group (OEWG).[3] Then in October 2020, a few months before the end of these two processes, France and Egypt, together with thirty-eight countries and the European Union, proposed the launch of a program of action for advancing responsible state behavior in cyberspace,[4] while two new resolutions were once again adopted by the UN General Assembly.[5]

At first sight, this profusion of initiatives looks like a renewed and strong interest among states in advancing cyber peace and stability. But the details reveal a more complex – and confusing – picture. Competing processes with overlapping mandates and agendas reflect the heightened strategic competition that prevails between great powers that pursue somewhat conflicting goals: Minimizing the risks to international peace, security, and cyber stability while maximizing their own cyber power, security, and normative influence. In other words, the cyber arms race is on and even though states aim at preserving collective security they are not ready

[1] *Paris Call for Trust and Security in Cyberspace*. (2018, November 12). Paris Call. https://pariscall.international/en/.

[2] UNGA Res. 73/266 (Dec. 5, 2018).

[3] UNGA Res. 73/27 (Dec. 22, 2018).

[4] *The Future of Discussions on ICTs and Cyberspace at the UN*. (2020, October 30). UNARM. https://front.un-arm.org/wp-content/uploads/2020/10/joint-contribution-poa-the-future-of-cyber-discussions-at-the-un-10302020.pdf.

[5] UNGA Res. 75/240 (Dec. 31, 2020).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

to give up any of their ability to conduct offensive operations in cyberspace.[6] The road to cyber peace is paved with malicious intentions.

This chapter offers an analysis of the multilateral efforts conducted over the past decade to build cyber peace in a context of proliferation of cyber conflicts and exacerbated geopolitical tensions, not to mention the global COVID-19 pandemic that has largely disrupted international meetings. It studies more specifically how international law has been leveraged in UN negotiations to serve strategic objectives. International law plays a central role in state-level discussions on peace and stability in cyberspace, but it has been a source of tension since the very first resolution of the UNGA on the regulation of cyberspace in 1998. Although considerable progress has been made by previous GGEs – notably in 2013 and 2015 – in achieving consensus over the applicability of international law to cyberspace, fundamental disagreements persist that are grounded in conflicting geopolitical representations and interests.

States not only have opposing views on the necessary means to ensure security and stability in cyberspace, but also on the content of the negotiations themselves. This reflects their diverging perceptions of the risks associated with the militarization of cyberspace and with the possible forms of responses authorized by international law in reaction to internationally wrongful acts. It also reflects the entanglement of the issues at stake: Negotiating on protective principles, such as the principle of sovereignty, for example, which may limit states' actions on the territory of other states, bears potential consequences that could extend to the lawfulness of the collection of transborder evidence.[7]

The first part of the chapter explains the context in which the two competing 2018 UN processes were created and, second, examines the challenging – and largely overlapping – mandates they were given. It then analyzes the October 2020 state initiatives as a window into the geopolitical underpinnings of cyber peace building going forward.

1 THE SHORT HISTORY OF CYBER PEACE BUILDING

The OEWG and the sixth GGE were created by resolutions 73/27 and 73/266, adopted within a few days, on December 5 and 22, 2018, respectively, in a context of heightened tensions between states. For the first time since the discussion started in 1998, two resolutions on ICTs in the context of international security – instead of one – were adopted by the General Assembly. While their composition and calendar differ, their mandates are largely similar, making them competing processes in essence. This situation testified to an apparent division between two blocks of member states opposing each other on this topic.

[6] Douzet, F. (2020), *Cyberspace: the New Frontier of State Power.* In Moisio S. et al. (Eds.), *Handbook on the Changing Geographies of the State: New spaces of geopolitics* (pp. 325–338), Cheltenham, UK: Edward Elgar.

[7] Delerue, F., Douzet, F. & Géry A. (2020), *The Geopolitical Representations of International Law in the International Negotiations on the Security and Stability of Cyberspac*e, IRSEM/EU Cyber Direct, pp. 50–55.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Their creation followed a series of preceding GGEs and of UN-level discussions on progress in information and telecommunication in the context of international security that reached a dead-end in June 2017 with the failure of the fifth GGE, triggering a series of private sector and multistakeholder initiatives to maintain international discussions on the security and stability of cyberspace.

The history of cyber peace building is still young but its analysis helps to measure the progress that has been made so far, and the scope of what remains to be done.

1.1 *How Cyberspace Became an International Security Issue in Multilateral Negotiations*

In 1998, the Russian Federation introduced the theme of "Progress in information and telecommunication in the context of international security" at the United Nations General Assembly, initiating a multilateral discussion on the consequences of the development of state and nonstate actors' cyber capacities on international security and stability (UNGA, Report of the First Committee, A/53/576 (1998)). This initiative led to the adoption of resolution 53/70 on December 4, 1998, by the General Assembly, which has since passed a resolution on the matter every year.

These resolutions created five successive GGEs up to 2016 (2004, 2009, 2012, 2014, and 2016). But the participants in the first GGE in 2004 proved unable to reach a consensus on a final report. As one of the experts in the Russian delegation later testified: "whether humanitarian international law and international law provided a sufficient regulation of security in international relations in case of a 'hostile' use of information and communication technologies for politico-military reasons was the main stumbling block."[8] Hence, international law was, from the start, at the heart of the disagreements among governmental experts.

The following three GGEs, however, were successful and led to the adoption of consensual reports in 2010,[9] 2013[10] and 2015[11]. These reports were submitted to the General Assembly by the Secretary General. The UNGA took note of the reports and suggested that member states draw from them.[12] The GGE reports contain recommendations on confidence building measures prone to preserve the security

[8] Streltsov, A. A. (2007), *International information security: description and legal aspects. ICTs and International Security.* Disarmament Forum, p. 8.

[9] UNGA, *Report of the Group of Governmental Experts on developments in the field of information and telecommunications in the context of international security*, A/65/201 (2010).

[10] UNGA, *Report of the Group of Governmental Experts on developments in the field of information and telecommunications in the context of international security*, A/68/98 (2013).

[11] UNGA, *Report of the Group of Governmental Experts on developments in the field of information and telecommunications in the context of international security*, A/70/174 (2015).

[12] UNGA Res. 65/41 (Dec. 8, 2010); UNGA Res. 68/243 (Dec. 27, 2013); UNGA Res. 70/237 (Dec. 23, 2015).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

and stability of cyberspace, along with measures of international cooperation and assistance that could be implemented by the states and, most importantly, norms of responsible state behavior in cyberspace.

The first major breakthrough was the recognition of the applicability of international law to cyberspace in the 2013 final report:

> International law, and in particular the Charter of the United Nations, is applicable and is essential to maintaining peace and stability and promoting an open, secure, peaceful, and accessible ICT environment.[13]

As a result, the following GGE was, for the first time, instructed to deal with international law.[14] Its final report in 2015 dedicated a full section (part 6) to international law, listing several rules. Since then, numerous states have endorsed this approach in their voluntary contributions to the Secretary General of the United Nations.[15]

The fifth GGE, however, ended in failure in June 2017, amid a dispute over the interpretation of international law. The governmental experts were indeed not able to reach an agreement for the adoption of a consensual final report. Three states – China, Cuba, and Russia – refused the explicit mention in the final report of the applicability of certain branches of international law, namely, the right of self-defense, the law of countermeasures, and the law of armed conflict. Cuban and Russian governmental experts explained that the endorsement of the applicability of these branches of international law in cyberspace could serve to justify the militarization of cyberspace,[16] and they pointed at profound divergences in interpreting the law. This mention was regarded as crucial by other states, particularly the United States, which released an unusually bitter communiqué blaming "some

[13] UNGA, *Report of the Group of Governmental Experts on developments in the field of information and telecommunications in the context of international security*, A/68/98, at ¶ 19 (2013).

[14] UNGA Res. 68/243 (Dec. 27, 2018).

[15] UNGA, *Developments in the field of information and telecommunications in the context of international security. Report of the Secretary General*, A/68/156/Add.1 (2013); UNGA, *Developments in the field of information and telecommunications in the context of international security: Report of the Secretary General*, A/69/112 (2014); UNGA, *Developments in the field of information and telecommunications in the context of international security: Report of the Secretary General*, A/69/112/Add.1 (2014).

[16] Representaciones Diplomáticas de Cuba en El Exterior (2017, June 23), 71 UNGA: *Cuba at the final session of Group of Governmental Experts on developments in the field of information and telecommunications in the context of international security*. http://misiones.minrex.gob.cu/en/un/statements/71-unga-cuba-final-session-group-governmental-experts-developments-field-information; Ministry of Foreign Affairs of the Russian Federation. (2017, June 29). *Response of the Special Representative of the President of the Russian Federation for International Cooperation on Information Security Andrey Krutskikh to TASS' Question Concerning the State of International Dialogue in this Sphere, Ministry of Foreign Affairs of the Russian Federation*. www.mid.ru/en/main_en/-/asset_publisher/G51iJnfMMNKX/content/id/2804288.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

participants" for the failure of the negotiations.[17] The representative of the United States was adamant:

> I am coming to the unfortunate conclusion that those who are unwilling to affirm the applicability of these international legal rules and principles believe their States are free to act in or through cyberspace to achieve their political ends with no limits or constraints on their actions. That is a dangerous and unsupportable view, and it is one that I unequivocally reject.[18]

The deadlock led a number of diplomats to claim that China and Russia were back tracking on the applicability of international law to cyberspace – which both countries denied – and that the discussion should continue among like-minded countries. The dreary perspectives over international discussions encouraged non-state actors to jump in, given the explosion of confrontation in cyberspace and its increasingly damaging consequences.

1.2 *A Multistakeholder Push to Reign in State Behavior*

The Snowden revelations in 2013 uncovered the extent of state offensive activities in cyberspace and made the security and stability of cyberspace a widely public and highly political issue, provoking the first summit bringing together the Internet governance community with the international security community: The so-called Net Mundial conference in 2014. The conference produced a statement with recommendations on Internet governance principles and a roadmap for the future evolution of the Internet governance ecosystem. This non-binding document was "the outcome of a bottom-up, open, and participatory process involving thousands of people from governments, the private sector, civil society, the technical community, and academia from around the world."[19] Since then, the proliferation of state-sponsored attacks started to backfire with large-scale consequences, undermining the security and stability of cyberspace for all users.

The private sector, academic actors, and other stakeholders who participate in Internet governance instances started to claim their own legitimacy and interest in taking part in the discussions over the security and stability of cyberspace. Academics created and built the Internet, later globalized and commercialized by the private sector. Most of the infrastructures are owned by major private companies that are at

[17] Markoff, M. G. (2017, June 23). *Explanation of Position at the Conclusion of the 2016–2017 UN Group of Governmental Experts (GGE) on Developments in the Field of Information and Telecommunications in the Context of International Security*. https://usun.usmission.gov/explanation-of-position-at-the-conclusion-of-the-2016-2017-un-group-of-governmental-experts-gge-on-developments-in-the-field-of-information-and-tele/.

[18] Ibid.

[19] NETMundial Multistakeholder Statement, April 24, 2014. https://netmundial.br/wp-content/uploads/2014/04/NETmundial-Multistakeholder-Document.pdf.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

the forefront of the attacks, often playing the role of first defender. Because of their data, resources, and skills, they are an essential partner of states for their cybersecurity. Global technology companies also have a vested interest in the security and stability of cyberspace for the trust of their users and the performance of their products, which are under constant attack.

Microsoft Corporation is by far the most important private actor in cybersecurity policymaking efforts, and leads multiple initiatives to promote cyber norms. As early as 2015, the company called on states – then on private companies – to adopt new norms. Most importantly, in 2017, its president, Brad Smith, proposed a Geneva Digital Convention for states to commit to protecting civilians against state-sponsored attacks, and the creation of an international organization for the attribution of cyberattacks.[20] The reference to international humanitarian law indirectly acknowledged the representation of cyberspace as a warfighting domain, but put the emphasis on the risk borne by civilians. The propositions were, however, regarded as infringing on states' rights and privileges. They were also criticized for shifting all the responsibility on states while creating few constraints on the industry to secure its products, whose flaws are exploited by malicious actors to conduct offensive operations.

The company then shifted its focus to promote cyber peace through multiple initiatives: A public petition, a commitment for the industry (Cybersecurity Tech Accord[21]), and the launch of the Cyberpeace Institute,[22] in partnership with the Hewlett Foundation and Mastercard in 2019. Its missions are to promote transparency and accountability by investigating and analyzing cyberattacks that impact civilians, provide assistance to the most vulnerable victims of cyberattacks, and promote cybersecurity norms of responsible behavior. The keyword is accountability, reflecting an interest in emphasizing state responsibility for the lack of cybersecurity. Other private sector initiatives were launched, such as the Charter of Trust,[23] initiated by Siemens in 2018, which contains ten principles to increase the resilience of digital products and the integrity of the supply chain.

The deadlock among states prompted the creation, in February 2017, of the Global Commission on the Stability of Cyberspace (GCSC), a multistakeholder group of international experts coming from academia, civil society and technical organizations, government, and the private sector. The Commission, initiated by the Ministry of Foreign Affairs of the Netherlands, and supported by several governments, private companies, and public organizations, started its work "convinced that an issue traditionally reserved to states—international peace and security—could no longer be addressed without engaging other stakeholders."[24] During its three-year

[20] Smith, B. (2017, February 14). *The Need for a Digital Geneva Convention*. Microsoft. https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/.

[21] Cyber Tech Accord. https://cybertechaccord.org/.

[22] Cyber Peace Institute. https://cyberpeaceinstitute.org/.

[23] Charter of Trust. www.charteroftrust.com/.

[24] Global Commission on the Stability of Cyberspace. https://cyberstability.org/about/.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

mandate, its mission was to propose norms and initiatives to guide responsible state and nonstate behavior in cyberspace in order to enhance international peace and security, with a main focus on stability, defined as such in its final report:

> Stability of cyberspace means everyone can be reasonably confident in their ability to use cyberspace safely and securely, where the availability and integrity of services and information provided in and through cyberspace are generally assured where change is managed in relative peace, and where tensions are resolved in a non-escalatory manner.[25]

In November 2017, the Global Commission proposed a Call to Protect the Public Core of the Internet, and that proposition has since been included in the European Union Cyber Security Act. It released its final report at the Paris Peace Forum of 2018 and the Internet Governance Forum held at the same time in Paris.

On the same occasion, the president of France launched the Paris Call for Trust and Security in Cyberspace (Paris Call, 2018), an initiative strongly supported by Microsoft, which led to a commitment to a set of principles and norms of responsible behavior of over 1,100 signatories, including 79 states, as of March 2021 – but not Russia, China, or the United States. The Paris Call refers to five GCSC norms, making explicit reference to three of them.[26] This initiative also demonstrates how some states attempt to draw from the legitimacy of multistakeholder support in order to build consensus over norms of responsible behavior for states and industry in cyberspace. This was also the approach favored by the Secretary General of the United Nations when setting up a High-Level Panel on Digital Cooperation in July 2018 to "advance proposals to strengthen cooperation in the digital space among Governments, the private sector, civil society, international organisations, academia, the technical community and other relevant stakeholders."[27]

Although states widely recognize the role of the private sector in the security and stability of cyberspace, and many of them endorse the multistakeholder governance model, they also perceive cyberspace as an international security threat that should be addressed by international regulation, which is the sole prerogative of UN Member States. It is in a very tense geopolitical context, marked by large-scale devastating attacks, information warfare targeting democratic processes, and the weakening of multilateral institutions that, eventually, the OEWG and the sixth UN GGE were created.

[25] Global Commission on the Stability of Cyberspace. (2019). *Advancing Cyberstability: Final Report*, p. 13.

[26] The Paris Call for Trust and Security in Cyberspace includes references to the norm on the public core of the Internet (Principle 2), the norm on the protection of electoral infrastructures (Principle 3), and the norm on hack back (Principle 8).

[27] U.N. Secretary General. (June 2019). *The Age of Digital Interdependence, Report of the UN Secretary General's High-level Panel on Digital Cooperation*, p. 39. Digital Cooperation. https://digitalcooperation.org/wp-content/uploads/2019/06/DigitalCooperation-report-web-FINAL-1.pdf.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

2 THE CREATION OF TWO COMPETING PROCESSES AT THE UN: THE OPEN-ENDED WORKING GROUP AND THE SIXTH GROUP OF GOVERNMENTAL EXPERTS

2.1 *A Context of Heightened Strategic Competition*

The resolutions creating the OEWG and the GGE were introduced by two groups of states, one led by the Russian Federation, the other one by the United States, forming seemingly adversarial blocs. But the reality is more complex and nuanced.

Russia, supported by China and other states,[28] proposed a first draft resolution in October 2018 creating an OEWG. The draft resolution listed not only norms adopted by the GGE in 2015, but also norms taken from the International Code of Conduct for Information Security proposed by the member states of the Shanghai Cooperation Organization in 2015 – and rejected by Western governments. In response, the United States submitted an alternative draft for a resolution creating a sixth GGE, which was supported by many European countries.[29] Eventually, Russia and cosponsoring states modified their project to account for the many criticisms they had received. But the United States and their cosponsors did not retract their own draft, arguing that the revised Russian draft still contained unacceptable provisions and did not reflect the 2015 GGE final report as well as it claimed. As a result, two competing resolutions on ICTs in the context of international security were debated in the First Committee of the UNGA; one promoted by Russia, the other by the United States. Both were adopted within a few days of each other, to the surprise of a number of states.

Heightened tensions between states surrounded the debates. According to the press communiqué describing the debates, Iran "[a]s a victim of cyber weapons," supported the "establishment of international legal norms and rules aimed at preventing the malicious use of cyberspace and information and communications technology" and condemned "those seeking dominance and superiority in cyberspace and their attempts to maintain the status quo" and pointed to a certain state (the United States) which, "in collaboration with Israel, used the computer worm

[28] Algeria, Angola, Azerbaijan, Belarus, Bolivia, Burundi, Cambodia, China, Cuba, Eritrea, the Russian Federation, Kazakhstan, Madagascar, Malawi, Namibia, Nepal, Nicaragua, Uzbekistan, Pakistan, the Syrian Arab Republic, the Democratic Republic of Congo, Samoa, Sierra Leone, Surinam, Tajikistan, Turkmenistan, Venezuela, and Zimbabwe. UNGA: *Developments in the field of information and telecommunications in the context of international security*, A/C.1/73/L.27/Rev.1 (2018).

[29] Germany, Australia, Austria, Belgium, Bulgaria, Canada, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Georgia, Greece, Hungary, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Malawi, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Ukraine, the United Kingdom, and the United States of America. UNGA: *Advancing responsible State behavior in cyberspace in the context of international security*, A/C.1/73/L.37 (2018).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Stuxnet against Iran's critical infrastructure, and yet has tabled a draft resolution regarding responsible state behaviour in cyberspace."[30]

The representative of China asked whether a negative vote on the Russian resolution would bring a "ticket" for the country to take part in the GGE, knowing that the number of participants is limited to twenty-five states, including the five permanent members of the UN Security Council.[31]

The debates gave the impression of two competing blocs of states, sponsoring different resolutions initiated by two states with diametrically opposed approaches on how to regulate cyberspace and what the content of the negotiations should be: On the one side, the United States and European countries, usually described as the "like-minded state," and on the other side, China and Russia. However, greater nuance is needed both in the homogeneity of the two blocs of states and the antagonism underlying their respective positions.

First, the countries in each group are not really homogeneous, they share certain characteristics in their approach that are not completely alike. There are, for example, important divergences between the Chinese approach and the Russian one,[32] as well as between France and the United States.

Second, the majority of UN member states did not adhere to any of the two groups and felt caught in the middle without a full grasp of the stakes. This supports an argument for the idea of two poles instead of two blocs of states structuring in international negotiations. More importantly, the vast majority of the member states voted in favor of both resolutions, as they regarded them as potentially complementary.[33] While these two processes might effectively be competing, they each advanced different sets of interests. The OEWG is open to all the member states, taking all the points of view into account. But, on the contrary, the composition of the GGE is limited to twenty-five member states designated "on the basis of equitable geographical distribution,"[34] the permanent members of the Security Council being *ex officio* members. Hence, the GGE appears as a more specialized entity

[30] Meetings Coverage, UNGA, First Committee Delegates Exchange Views on Best Tools for Shielding Cyberspace from Global Security Threats Triggered by Dual-Use Technologies, GA/DIS/3613 (Oct. 30, 2018).

[31] Meetings Coverage, "First Committee Approves 27 Texts, Including 2 Proposing New Groups to Develop Rules for States on Responsible Cyberspace Conduct, Meetings Coverage," GA/DIS/3619 (Nov. 8, 2018).

[32] Broeders, D., Adamson, L. & Creemers, R. (2019, November 5). A *Coalition of the Unwilling? Chinese and Russian Perspectives on Cyberspace*. Universiteit Lieden. www.universiteitleiden.nl/en/research/research-output/governance-and-global-affairs/a-coalition-of-the-unwilling-chinese-and-russian-perspectives-on-cyberspace.

[33] The resolution "Developments in the field of information and telecommunications in the context of international security" (UNGA Res. 73/27 (Dec. 5, 2018)) was adopted with 119 votes against 46 and 14 abstentions (UNGA A/73/PV.45 (2018)) and the resolution "Advancing responsible State behaviour in cyberspace in the context of international security" (UNGA Res. 73/266 (Dec. 22, 2018)) was adopted with 138 votes against 12 and 16 abstentions (UNGA A/73/PV.65 (2018).

[34] UNGA Res. 73/266, ¶ 3 (Jan. 2, 2019).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

which could lead to concrete progress on the core questions debated, whereas the nonlimited composition of the OEWG offers a more inclusive approach that allows each state to have its positions and interests heard.

The first session of the OEWG, which took place in New York in September 2019, actually highlighted the interests that many states have in taking an active part in the discussions – something confirmed by the high number of states involved in the second formal session in February 2020, as observed through the online videos of the debates on the UN website. Hence, the two ongoing processes are somewhat complementary. Despite the hostile climate that surrounded their creation, which reveal strong geopolitical tensions, they offer – in theory at least – a possibility for states to go beyond their inherent divisions and offer a smooth parallel functioning, or even synergy. The ambassadors Guilherme de Aguiar Patriota and Jürg Lauber, who preside over the GGE and the OEWG, respectively, actually advertised this constructive ambition from the moment they were nominated in these roles, as they have publicly declared on multiple occasions.

The complementarity of the two cyber norms processes has been highlighted by several states. However, an analysis of their respective mandates shows that, if they can be complementary, their mandates overlap to a certain extent, which does not facilitate the search for consensus and coherence in the negotiations.

2.2 *Overlapping Mandates and Subtle Differences*

At first glance, the mandates of the two groups are so similar they overlap to a large extent, with the risk of encroaching on one another. Indeed, both groups are mandated to work on the norms, rules, and principles of responsible behavior of the states, on confidence building measures, on capacity building, and international law. However, a careful reading reveals several differences.

First, the GGE can consult states that are not part of the GGE and competent regional organizations such as the African Union, the Organization of American States, the Organization for Security and Co-operation in Europe, and the Regional Forum of the Association of Southeast Asian Nations. The OEWG, on the other hand, is empowered to hold informal sessions to consult private actors and nongovernmental organizations. Furthermore, nonstate actors are authorized to attend the formal sessions as long as they have an accreditation with the United Nations Economic and Social Council (ECOSOC), following the Chinese refusal to further enlarge the pool.

Second, the GGE report is to be presented to the General Assembly with "an annex containing national contributions of participating governmental experts on the subject of how international law applies to the use of information and communications technologies by states."[35] As such, the twenty-five countries participating

[35] UNGA Res. 73/266, ¶ 3 (Jan. 2, 2019).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

in the GGE will have to clarify their position on the international law applicable to cyber operations. Some states, such as France and the Netherlands, have already moved forward in this regard. The French Ministry of Armed Forces published a report, *International Law Applied to Cyberoperations*,[36] in 2019, and the Dutch Ministry of Foreign Affairs also published *International Law in Cyberspace* in 2019.[37] These documents are most likely meant to be the two countries' national contributions to the GGE.[38]

Finally, the OEWG is tasked with examining "the possibility of establishing regular institutional dialogue with broad participation under the auspices of the United Nations"[39] to deal with ICTs in the context of international security. It could take the form of a permanent body or a new process.

A number of differences have raised concerns, starting with the respective timelines. The OEWG was supposed to end its work in 2020 and submit its report to the UNGA during its 75th session, a year before the GGE. Indeed, the GGE's mandate ends in May 2021 and the GGE should thus present its report to the UNGA during its 76th session. The extension of the 75th session until March 2021, due to the COVID-19 crisis, allowed the OEWG's work to continue in order to present it to the 76th session of the UNGA. The final deadlines for the two reports have therefore been preserved. Yet, some observers worry that several states behind the resolution creating the OEWG might change course after the end of its sessions. In other words, they would be adopting a constructive approach up to the end of the OEWG's work in order to achieve a consensus on its conclusions, before becoming less cooperative during the remaining time of the GGE sessions to push for a failure, and boast of the superior achievements of the OEWG. But given the short time between the end of the two processes, this might be more difficult to achieve.

The second concern regards the content of the mandates. Both processes discuss international law, which constitutes a central topic in their proceedings. This can be seen both as an opportunity and a risk: States may conduct meaningful discussions and make progress on a consensus about the interpretation of international law in this new context of international peace and security, but they also may take

[36] France, Ministry of Armed Forces. (2019, September 9). *International law applied to cyberoperations.* www.defense.gouv.fr/content/download/565895/9750877/file/Droit+internat+appliqu%C3%A9+aux+op%C3%A9rations+Cyberespace.pdf.

[37] Netherlands (made public on September 26, 2019). *Letter of July 5, 2019 from the Minister of Foreign Affairs to the President of the House of Representatives on the international legal order in cyberspace. Annex.* www.government.nl/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace.

[38] For a compared study of the states' positions on international law applied to cyberoperations, see Roguski, P. (2020). *Application of International Law to Cyber Operations: A Comparative Analysis of States' Views.* The Hague Program on Cyber Norms. www.thehaguecybernorms.nl/research-and-publication-posts/application-of-international-law-to-cyber-operations-a-comparative-analysis-of-states-views.

[39] UNGA Res. /27, ¶ 5 (Dec. 5, 2019).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

diverging directions in the two processes, leading to a certain level of instability for the international legal order.

This concern also applies to norms of responsible state behavior, mentioned twice in resolution 73/27 that defines the mandate of the OEWG. The situation here is delicate for two reasons. The first mention of norms in resolution 73/27 appears early on in the definition of the OEWG mandate in paragraph 5.[40] Norms – as stated in the resolution – constitute the working base of the OEWG, but their definition is slightly different from the norms of the 2015 GGE report to which they refer. The mandate of the GGE is clearer since resolution 73/266 refers exclusively to the GGE report. As a result, the working base of the two processes could slightly differ and potentially increase the risks of divergence, or even contradiction in the meaning of the recommendations adopted by each process. For example, the recommendation on the prevention of malicious computer tools or technologies is included in a paragraph on supply chain integrity in the 2015 GGE report, whereas it is the subject of a stand-alone provision in resolution 73/27 that creates the OEWG. This could indicate a desire to work more extensively on the issue of proliferation in the context of the OEWG.

The practice of the states, however, shows that this risk remains limited as a large majority of states, during the first two sessions of the OEWG, opted for the norms as stated in the 2015 GGE report. This illustrates the lack of consensus on the norms as stated in the provisions of resolution 73/27, but it also highlights a gap between a strict application of the mandate and the practice adopted by states during the negotiations.

The uncertainty around the working base could also affect other aspects of the negotiations, such as norm implementation.[41] Member states are tasked with detailing the operationalization of the norms. Because several of them are quite vague, they need to be specified in order to be implemented. Finally, the OEWG mandate paves the way for a possible reappraisal of the agreed provisions of the 2013 and 2015 GGEs as states are able to "introduce changes,"[42] including establishing new norms. Elaborating new norms is authorized by resolution 73/27 and could involve creating new norms that better define what responsible behavior is, or revisit the norms adopted in the 2013 and 2015 reports.

The second mention of norms in the resolution 73/27 can be found in the second part of the definition of the mandate. But it does not state explicitly if this mention refers to the norms stated in resolution 73/27 or the ones adopted by the GGEs in 2013 and 2015.

[40] "[A]cting on a consensus basis, to continue, as a priority, to further develop the rules, norms and principles of responsible behaviour of states listed in paragraph 1 above, and the ways for their implementation; if necessary, to introduce changes to them or elaborate additional rules of behaviour." UNGA Res. /27, ¶ 5 (Dec. 5, 2019).

[41] UNGA Res. 73/27, ¶ 5 (Dec. 5, 2019).

[42] Ibid.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

A close reading of the mandate thus highlights a number of questions related to the working base on which the negotiations are to be conducted. The practice of using the GGE norms have prevailed so far, but contradictions could emerge as both the GGE and the OEWG are tasked with working on these provisions.

It was also hard to know how the work would be divided between the two processes, given the fact that international law and norms of responsible behavior are mentioned in both mandates. In his speech during the first session of the OEWG in June 2019, the special representative of the President of the Russian Federation for international cooperation in information security proposed that the OEWG deals with norms of responsible behavior, confidence building measures, and measures of international cooperation and assistance, hence leaving the issue of international law to the GGE.[43] This proposal was not accepted. As a result, both processes work concomitantly on the entire set of issues.

This situation is both understandable and problematic. On the one hand, international laws and norms of responsible state behavior are intrinsically linked and, therefore, difficult to completely dissociate. On the other hand, this situation reinforces the risk of repetitions in the content of the negotiations, and also the risk of contradictions in the recommendations made by the two groups on the rights and obligations of states. Most importantly, the refusal to dissociate them highlights disagreements on the necessary means to ensure security and stability of cyberspace.

The COVID-19 pandemic has added a layer of complexity. In addition to overlapping mandates, the two processes have ended up with largely overlapping calendars since the two final reports will be produced a month apart from each other. It is, however, difficult to assess whether this overlapping can help build synergy between the two processes or fuel further rivalry. Most importantly, states have not waited for the end of these two processes, as initially planned, to propose new processes.

3 BUMPY ROAD TO CYBER PEACE

3.1 *New Path(s) for Cyber Stability?*

In the face of potential difficulties in reaching consensus over a final report and successfully coordinating the two existing processes, France and Egypt, supported by thirty-eight countries and the European Union, proposed on October 1, 2020, a new path to cyber stability: The creation of a Program of Action (PoA) for advancing responsible state behavior in cyberspace, a proposal made to all member states

[43] Embassy of the Russian Federation to the United Kingdom of Great Britain and Northern Ireland (June 7, 2019). *Statement by Amb. Andrey Krutskikh, Special Representative of the President of the Russian Federation for International Cooperation in the Field of Information Security at the First Session of the UN Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security, New York, 3–4 June 2019*. https://rusemb.org.uk/article/541.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

within the context of the OEWG. Neither China, Russia, nor the United States have been officially part of this initiative.

A PoA consists of the production of an outcome document adopted by an intergovernmental conference, considered as politically binding, which contains objectives, recommendations, and rules for implementation and monitoring, in a new process with working conferences every other year including a review conference every five years.[44] It would, therefore, fulfill one of the objectives of the OEWG; that is, "study the possibility of establishing a regular institutional dialogue with broad participation under the auspices of the United Nations."[45]

This process would present the advantage of bringing the discussion back into a single process more inclusive than the GGE. As a new process, it would also be free from all the political baggage linked to the United States versus Russia rivalry over the GGE and OEWG processes. Unlike previous dialogues, it would not require building a consensus over a final report but, rather, building a working relationship that fosters practical cooperation and allows for agreement on specific issues as the discussions progress. There would be no end dates, even if states fail to agree on an outcome document at the end of a technical or review conference. The ultimate goal is to preserve and build on the agreed provisions of the previous GGE by providing a "forum for practical cooperation and ongoing discussions."[46]

Although the proposition was well received, two draft resolutions were put forward before the UNGA First Committee a few days later.[47] On October 5, a coalition of forty-six member states led by the United States, including France and many supporters of the PoA, proposed a draft resolution entitled "Advancing responsible state behaviour in cyberspace in the context of international security." The resolution acknowledges the ongoing discussions at the GGE and OEWG and declares that member states will study the conclusions of both groups and "will decide thereafter on any future work, as needed."[48]

The very next day, jumping ahead of the calendar, Russia along with fourteen other states proposed another draft resolution stating – in operative paragraph 1 – that the UNGA will create a new OEWG starting in 2021, without waiting for the conclusions of the two ongoing processes.[49] A revised version was submitted

[44] Delerue, F. & Géry, A. (2020, October 6). *A New UN Path to Cyber Stability*. Directions Blog. https://directionsblog.eu/a-new-un-path-to-cyber-stability/.

[45] UNGA Res. 73/27, ¶ 5 (2018).

[46] Australia. (2020, December 2). *Informal Australian Research Paper: What Next for Advancing Responsible State Behaviour at the United Nations*. https://front.un-arm.org/wp-content/uploads/2020/12/australian-research-paper-revised-december-2020-version-2-oewg-regular-institutional-dialogue.pdf.

[47] UNGA, *Developments in the field of information and telecommunications in the context of international security*. Report of the First Committee, A/75/394 (2020).

[48] UNGA, *Advancing responsible state behavior in cyberspace in the context of international security*, A/C.1/75/L.4 (2020).

[49] UNGA, *Establishment of a nuclear-weapon-free zone in the region of the Middle East*, A/C.1/75/L.8 (2020).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

on October 26, specifying that the new OEWG "shall start its activities up to the conclusion of the work of the current Open-Ended Working Group and considering its outcomes."[50] The revised version, however, leaves room for interpretation as to whether the acquis will be preserved, since the mandate of the new OEWG includes the possibility to "if necessary, … introduce changes to them [the norms] or elaborate additional rules of behaviour."[51] In addition, this new draft resolution borrows from the PoA approach by stating that the new OEWG "may decide to establish thematic subgroups, as the Member States deem necessary, with a view to fulfilling its mandate and facilitating the exchange of views among States on specific issues related to its mandate, and may decide to interact, as appropriate, with other interested parties, including businesses, non-governmental organizations and academia."[52] Yet, while it opens the door to consultations with nonstate actors, the drafting is less prescriptive than in the resolution that created the first OEWG, and it will limit nonstate actors' participation in the discussions for the next five years. And, finally, there is a tweak that leaves the question of its future mandate open: The name changed from "OEWG on developments in the field of information and telecommunications in the context of international security" to "OEWG on security of and in the use of information and communication technologies."[53]

Both draft resolutions were submitted to a vote at the First Committee on November 9, 2020, and both were adopted. The UNGA adopted both of them respectively on December 7th (UNGA Res. 75/32 (2020)) and December 31st (UNGA Res. 75/240 (2020)), adding more confusion to the field of competing processes. The PoA was proposed to all participating states during the discussions held within the OEWG, and offered to continue the negotiations within a single process. The resolution sponsored by Russia offered to continue this dialogue within the OEWG and the resolution sponsored by the United States suggested to wait and see. These competing initiatives have fostered strong debates within the United Nations and, more broadly, among actors involved on these matters.

3.2 *The Contest for Normative Influence*

Once again, the debates seemed to oppose two blocs, one led by the Russian Federation and the other by Western states along with Australia, even though the reality was more complex. We studied the coalition of sponsors and the votes at the UNGA for each resolution. The analysis reveals that the United States gained support among states since its 2018 resolution, while Russia has lost part of its support (Figure 9.1).

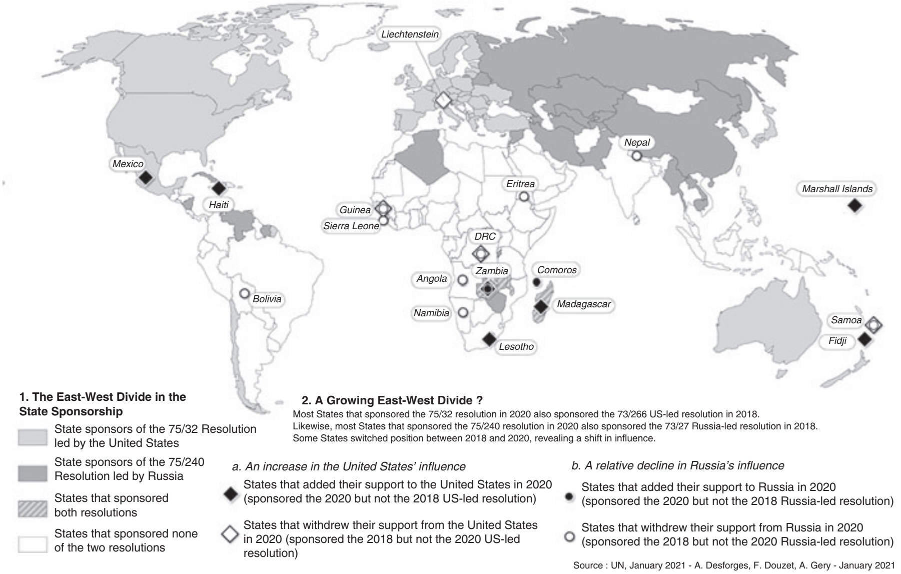
[50] UNGA, *Developments in the field of information and telecommunications in the context of international security*, A/C.1/75/L.8/Rev.1 (2020).

[51] Ibid., ¶ 1.

[52] Ibid., ¶ 4.

[53] Ibid., op. ¶ 4.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D


Liechtenstein
Mexico
Haiti
Nepal
Marshall Islands
Eritrea
Guinea
Sierra Leone
DRC
Comoros
Zambia
Angola
Madagascar
Bolivia
Namibia
Lesotho
Samoa
Fidji
1. The East-West Divide in the State Sponsorship
State sponsors of the 75/32 Resolution led by the United States
State sponsors of the 75/240 Resolution led by Russia
States that sponsored both resolutions
States that sponsored none of the two resolutions
2. A Growing East-West Divide ?
Most States that sponsored the 75/32 resolution in 2020 also sponsored the 73/266 US-led resolution in 2018.
Likewise, most States that sponsored the 75/240 resolution in 2020 also sponsored the 73/27 Russia-led resolution in 2018.
Some States switched position between 2018 and 2020, revealing a shift in influence.
a. An increase in the United States' influence
States that added their support to the United States in 2020 (sponsored the 2020 but not the 2018 US-led resolution)
States that withdrew their support from the United States in 2020 (sponsored the 2018 but not the 2020 US-led resolution)
b. A relative decline in Russia's influence
States that added their support to Russia in 2020 (sponsored the 2020 but not the 2018 Russia-led resolution)
States that withdrew their support from Russia in 2020 (sponsored the 2018 but not the 2020 Russia-led resolution)

Source : UN, January 2021 - A. Desforges, F. Douzet, A. Gery - January 2021

FIGURE 9.1 State sponsorship of 2020 UN Cyber Diplomacy Resolutions: a persistent east-west divide.

The map "State Sponsorship of 2020 UN Cyber Diplomacy Resolutions" illustrates a clear east-west divide regarding the sponsorship of the two resolutions. The US-led resolution 75/32 was overwhelmingly supported by Western countries while the Russian led resolution 75/240 was supported by Eastern Countries. But the map also reveals a slight change of balance in favor of the United States. In 2020, eight states that had sponsored the Russian-led resolution in 2018 withdrew their support to Russia for the 2020 resolution. In the meantime, two states (Comoros and Zambia) added their support to Russia; that is, sponsored the 2020 resolution but not the 2018. But Zambia also sponsored the US-led resolution. On the contrary, the US-led resolution gained sponsorship between 2018 and 2020: Seven states added their support to the United States in 2020 while four withdrew their support, as illustrated by the graph in Figure 9.2.

The two draft resolutions were introduced before the UN First Committee on the October 5–6, 2020. The first one, "Advancing responsible State behaviour in cyberspace in the context of international security,"[54] was introduced by the United States on behalf of fifty-three states, against fifty-one states for the 2018 US-sponsored resolution.[55] The vote at the First Committee reached a large

[54] UNGA, *Advancing responsible State behavior in cyberspace in the context of international security*, A/C.1/75/L.4 (2020).

[55] UNGA, *Developments in the field of information and telecommunications in the context of international security*. Report of the First Committee, A/73/505 (2018).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

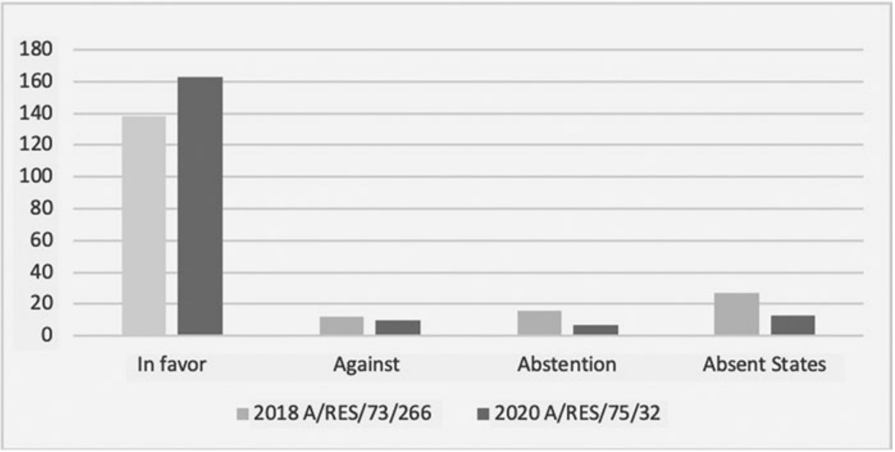


FIGURE 9.2 The 2020 US-led resolution gains more votes than the 2018 resolution.

consensus with 153 states in favor, 9 against, and 9 abstaining. The UNGA adopted the resolution in its plenary session on December 7, 2020, by an even larger margin: 163 in favor, 10 against, and 7 abstaining. By comparison, the 2018 US-sponsored resolution was adopted by a lower margin (138 in favor, 12 against, 9 abstaining). This can be explained by the noncontentious nature of the 2020 resolution, which did not involve a strong commitment to a specific process.

The draft resolution A/C.1/75/L.8/Rev.1, sponsored by Russia on behalf of twenty-six states (thirty-four in 2018), however, was faced with harsh criticism coming mainly from Western states. The representative of the Russian Federation, speaking in exercise of the right of reply, said: “Western delegations are sabotaging the process and breaking with decades of consensus on cybersecurity.” As such, his delegation was offended by their level of cynicism and hypocrisy, which stalled the work of the OEWG. He added, “If it were not for the Russian Federation, the United Nations would not have open negotiations on the matter.”[56]

The opposition focused on operative paragraph 1, creating a new OEWG for 2021. Western states objected that it is part of the mandate of the present OEWG to make suggestions about future institutional work and, therefore, decide whether a new OEWG should be created. The draft resolution would thus preempt the work

[56] Meeting’s coverage, UNGA (2020, November 9), First Committee Approves 15 Draft Resolutions, Decisions on Disarmament Measures, Including 2 Following Different Paths towards Keeping Cyberspace Safe, GA/DIS/3659 (Nov. 9, 2020).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

of the present OEWG. They therefore asked for the withdrawal of this operative paragraph and all related ones.

The Russian delegates strongly opposed this demand; they believed that this would void the resolution of all substance and invoked article 129 of the Rules of Procedures of the UNGA[57] to have the contentious operative paragraph 1 be voted on separately instead of withdrawn. This situation in itself illustrates the opposition between Western states and the Russian Federation. As a result, the President of the First Committee put to a vote the decision regarding the division of the draft resolution, which was approved by fifty-seven states in favor, thirty-one against, and sixty-three abstaining. Once the division approved, the First Committee then proceeded to the three following votes on: the preamble (108 in favor, 49 against, 11 abstaining); the operative paragraph 1 (92 in favor, 52 against, 24 abstaining); and the resolution as a whole (104 in favor, 50 against, 20 abstaining).

The resolution was thus submitted to the UNGA and adopted on December 31, 2020. The date in the middle of the holiday season may explain the high number of absent states on the day of the vote. The voting data show an overall support for the resolution and also a sizeable opposition: ninety-two in favor, fifty against, and twenty-one abstaining. The Russia sponsored resolution was nevertheless adopted by the UNGA, yet the number of States voting in favor (92) was drastically lower than for the 2018 Russia sponsored resolution (119 in favor). However, this result must be interpreted with caution. Thirty states were absent from the UNGA that day, among which eighteen states who voted in favor of the Russia sponsored resolution in 2018. A close reading of the votes shows, however, that Russia indeed lost the support of an additional thirteen member states compared to 2018, as illustrated by the graph in Figure 9.3.

The charts "The 2020 UNGA Balance of Votes" illustrate the percentage of states that voted in favor of each resolution, against it, or abstained (Figure 9.4).

The map "UNGA Vote on 2020 Cyber Diplomacy Resolutions," with the votes on the two resolutions, highlights the dynamics of power between states. First, it confirms the East-West divide observed on the state sponsorship map. It also confirms the growing support gained by the United States, whose resolution was adopted by a larger and growing margin of states (with fewer absent states) and by less opposition. In addition, support for the US-led resolution appeared more consistent. All the states that had only sponsored the US-led resolution in 2020 voted for it and, in

[57] *"A representative may move that parts of a proposal or of an amendment should be voted on separately. If objection is made to the request for division, the motion for division shall be voted upon. Permission to speak on the motion for division shall be given only to two speakers in favour and two speakers against. If the motion for division is carried, those parts of the proposal or of the amendment which are approved shall then be put to the vote as a whole. If all operative parts of the proposal or of the amendment have been rejected, the proposal or the amendment shall be considered to have been rejected as a whole."*

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

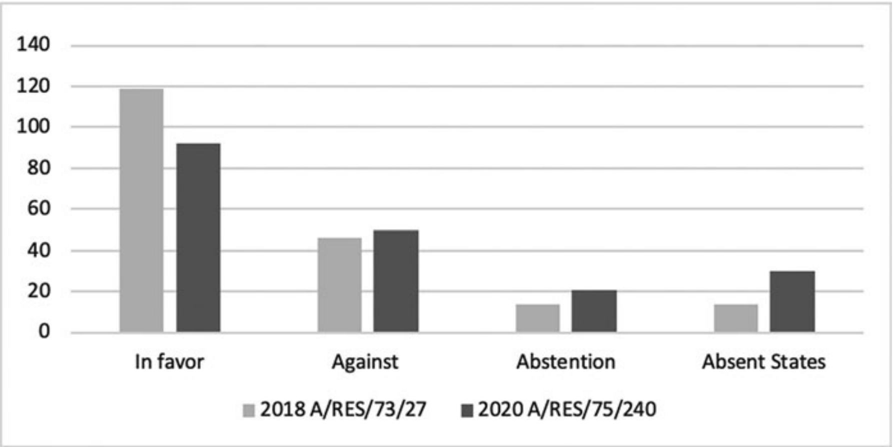


FIGURE 9.3 The 2020 Russian-led resolution gathers less votes than the 2018 resolution.

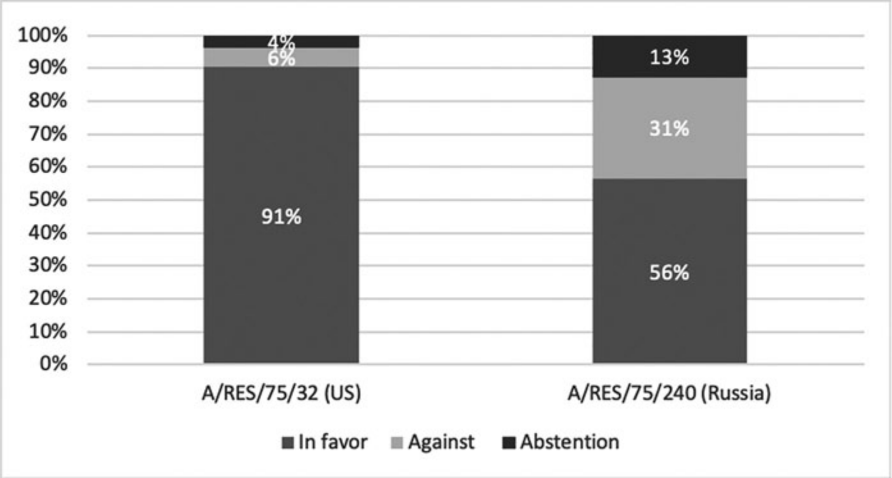
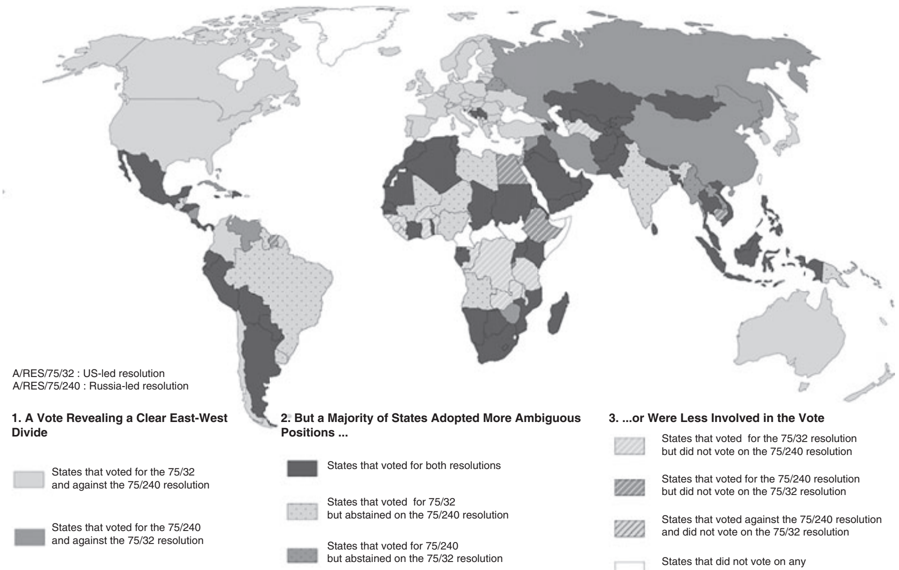


FIGURE 9.4 The 2020 UNGA balance of votes.

addition, voted against the Russia-led resolution (none of them abstained or voted in favor of it) (Figure 9.5).

On the contrary, several states that had sponsored the Russia-led resolution did not oppose the US-led resolution: They either voted in favor of it or abstained. This could be explained by the fact that the US-led resolution is more consensual than the Russia-led resolution, but it also reveals a more complex picture. A majority of states either voted for both resolutions or voted for one and abstained from

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D


A/RES/75/32 : US-led resolution
A/RES/75/240 : Russia-led resolution
1. A Vote Revealing a Clear East-West Divide
States that voted for the 75/32 and against the 75/240 resolution
States that voted for the 75/240 and against the 75/32 resolution
2. But a Majority of States Adopted More Ambiguous Positions ...
States that voted for both resolutions
States that voted for 75/32 but abstained on the 75/240 resolution
States that voted for 75/240 but abstained on the 75/32 resolution
3. ...or Were Less Involved in the Vote
States that voted for the 75/32 resolution but did not vote on the 75/240 resolution
States that voted for the 75/240 resolution but did not vote on the 75/32 resolution
States that voted against the 75/240 resolution and did not vote on the 75/32 resolution
States that did not vote on any of the two resolutions

Source : UN, January 2021 - A. Desforges, F. Douzet, A. Gery - January 2021

FIGURE 9.5 UNGA vote on 2020 Cyber Diplomacy Resolutions: a majority of states caught between two stools.

the other. This shows that the East-West divide is clear, but most states – caught between two stools – chose not to position themselves within this duopoly. Any claim that international negotiations on the security and stability of cyberspace is marked by a strong opposition between two blocks of states should thus be cautioned.

CONCLUSION

The cyber peace building dynamics at the United Nations reflects fundamental disagreements on the means to ensure the security and stability of cyberspace and the struggle for normative influence among states.

Russia has justified its 2020 initiative by the desire to ensure that international discussions would continue after the end of the two processes, highlighting its role in opening negotiations. But the Russian Federation might also be defending another agenda, along with its own legal culture and perspective. Russia makes no secret of wanting to elaborate a treaty for cyberspace, an option best preserved by the OEWG process. A PoA, on the contrary, could considerably delay the perspective of a treaty by providing a process with no end date and "politically binding" decisions, a compromise that is *a priori* at odds with Russia's legalist approach to international relations. Yet, Russia could also use the PoA as a vehicle to launch the drafting process of a treaty.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

The analysis of the maps shows there is a strong polarization between the United States and Russia and a relative decline in Russia's influence. However, Russia's leadership is still strong enough to get its resolution voted by the UNGA and there is still a vast reserve of votes, given the ambiguous position of a significant number of states. Indeed, a majority of states did vote for both resolutions, or chose to vote for one resolution without opposing the other.

To the surprise of all observers, states participating in the OEWG were able to reach a consensus and adopt a report on March 12, 2021,[58] while the GGE had still not ended its mandate. Meanwhile, a new OEWG is scheduled to start its work soon after the adoption of the consensus report since the UNGA enacted its creation in resolution 75/240. This leaves the question of the creation of other processes totally open, particularly since the PoA proposal has been acknowledged by the OEWG. Indeed, the final report recommended that "the Programme of Action should be further elaborated including at the Open-Ended Working Group process established pursuant to General Assembly resolution 75/240."[59] Although the report states that the PoA should be discussed within the future OEWG, it also leaves room for discussion of a PoA in another context. In this regard, the French Ambassador for Digital Affairs, Henri Verdier, announced on March 24, 2021 that France was considering launching the PoA in October 2021[60]; that is, at the beginning of the 76th session of the UN General Assembly. If this was to happen, it would raise the question of how many processes could states handle without ending in a total deadlock, letting alone the fact that another GGE could also be created in the meantime. While the PoA could offer a productive venue for states that wish to work on more action-oriented recommendations, it could also lead to more bumps in the road to cyber peace.

The road to cyber peace is arduous, given the will of states to preserve their ability to conduct cyber offensive operations. Official documents tend to refer to cyber stability rather than cyber peace as a goal for international negotiations.[61] The proliferation of damaging attacks and the risk of conflict escalation in cyberspace have led states to leverage the traditional instruments of collective security – such as international law and nonbinding norms of responsible behavior – to regulate cyberspace. In the early stages of consensus building up

[58] UNGA, *Final report of the OEWG*, A/AC.290/2021/CRP.2 (2021).

[59] Ibid., ¶ 77.

[60] Statement of the French Ambassador for Digital Affairs Henri Verdier at the launching meeting of the working group 3 of the Paris Call for Trust and Security in Cyberspace (March 24, 2021).

[61] The Global Commission has given its own definition of Stability of Cyberspace: "Stability of cyberspace means everyone can be reasonably confident in their ability to use cyberspace safely and securely, where the availability and integrity of services and information provided in and through cyberspace are generally assured, where change is managed in relative peace, and where tensions are resolved in a non-escalatory manner." Global Commission on the Stability of Cyberspace. (2019). *Advancing Cyberstability: Final Report*, p. 13.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

to 2016, these instruments have helped advance the discussions by providing an existing legal framework applicable to cyber operations as a basis for negotiation. But since then, the renewed strategic competition and exacerbated geopolitical tensions have led states to engage not only in a cyber arms race, but also in a competition for normative influence. As a result, international law has proved to be exactly what it is: An instrument in the service of state foreign policy – with the risk to lead states to a stalemate.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

PART IV

Reflections and Research Notes

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

10

Imagining Cyber Peace

An Interview with a Cyber Peace Pioneer

Camille François and Christopher Ankersen

CHRISTOPHER ANKERSEN: What, to you, is cyber peace?

CAMILLE FRANÇOIS: For me, cyber peace is the set of norms and behaviors that we want democratic societies to observe in cyberspace, both below and above the threshold of armed conflict. It is a recognition that when we think about how to deploy cyber power, you also have to take into account what does it mean for democracy? What does it mean for human rights? It's a positive framework that talks about how you want to behave, and what you want to preserve, as you're thinking through deployment of cyber power.

CHRISTOPHER ANKERSEN: I think it's very interesting that you've connected cyber peace to the idea of democracy. Do you think, therefore, that it's not possible for other kinds of countries to play a role in this? Are they always going to be the "others" in this exercise?

CAMILLE FRANÇOIS: When I started working on cyber peace, my focus was working on both the US and the French approaches to cyber power. I was looking through historical records of how cyber power was defined, and it was very evident that cyber power was only defined in the context of warfare and conflict. Similarly, it was very obvious that cyber warfare was defined without its companion question, which is what is cyber peace? I thought that this was backwards; I thought that it was important for democracies, who are thinking through what cyber power is and how to deploy it, to have a positive vision of cyber peace, and you deploy cyber power outside the realm of war which, again, was a clear gap.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

CHRISTOPHER ANKERSEN: It's very interesting to link it back to this idea of cyber power. Do you think then that cyber peace is a goal? What I mean is, countries are deploying cyber power in order to "do things." Is cyber peace, one of those things they're trying to do? Or do you think it's more like a precondition or even a collateral outcome?

CAMILLE FRANÇOIS: It's a necessary question for the societies to answer. Peace is a state of affairs that is much more common than war, which is what we want. And so it is interesting and somewhat baffling to me that most of the governments whose cyber theories we work on have spent all this time trying to work through the minutiae of how you deploy cyber power in wartime, which is important, but without ever touching on what the considerations are that you go through to get there. What are the appropriate sets of norms? How do you want to deploy cyber power in peacetime? And I think that this blind spot is detrimental to peace and stability.

When I started working on this, people were perhaps confused: it sounded like a "hippie" theory. But I think the past few years have demonstrated that the major cyber incidents do happen in peacetime and that the spectrum of conflict and conflict evolution doesn't allow these democratic societies to have a space for thinking how they deploy cyber power in peacetime. And this has to be a necessary democratic conversation.

CHRISTOPHER ANKERSEN: Can you go into a little bit more what you mean by people's reactions to cyber peace?

CAMILLE FRANÇOIS: So from a research perspective, I was looking at two bodies of conceptions on the role of the state in cyberspace. The first body of work that I was looking at was the cyber utopians. (It's the John Perry Barlow school of thought, to be brief.) And that's a really interesting body of work because, initially, it conceptually makes no room for state cyber power. The essence of the declaration says "you giants of flesh and steel have no room where we gather." A conception of cyberspace that makes no room for the deployment of state cyber power. And that's interesting. But it creates a huge gap between where we are and that initial conception. That

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

body of work is preoccupied with cyber security. It's also a school of thought that has thought a lot about encryption, but it kind of stops at actual cyber power. Because, again, it conceptually doesn't make room for that.

My other point of departure was actual military cyber theory, which is almost the radical opposite of where the cyber utopians are starting from. In it, cyber power deploys itself all over cyberspace, regardless of where we are on the spectrum of conflict and peace on wartime.

And so, looking at these two bodies of work, one says state cyber power is nowhere. The other one says state cyber power is everywhere. And for me it was self-evident that we were lacking the sort of rational approach that says today we are in a situation where states are building cyber power, they are building sort of military theories on how to express cyber power in cyberspace, and we need to have the in-between conversation, which is: What is the desirable use of that power? What is the responsible use? What is the democratic use of cyber power in peacetime?

And that was my point of departure – being stuck in between these two bodies of work, seeing the obvious gap, the conversation that has not happened.

CHRISTOPHER ANKERSEN: It's quite fascinating that the utopians saw cyberspace as almost anarchic, in a libertarian sense, where everything was possible. And we see this crop up over and over again: With the advent of social media, we had the same optimism. "Oh, great! Tahrir Square, uprisings across the Arab Spring, now we will know exactly what's going on. We won't have to worry about things being mediated!" But it really only took one contact with reality to see that wasn't exactly the case. Do you think, therefore, that this idea of cyber stability (as opposed to cyber peace) is a compromise, a way of trying to avoid the disappointment experienced before? Along the lines of "Well, let's not worry about peace, but can we at least have some kind of rules of the road so that we can have some reliability?" Do you think that stability is an ingredient towards cyber peace? Or is it a completely different approach?

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

CAMILLE FRANÇOIS: So it's a really interesting question, because one of the things I was circling around while working on cyber peace was also the question of what type of entities belong at the table when we talk about the reasonable deployment of cyber power in peacetime. When I started this body of research, I was at the Berkman Center for Internet and Society at Harvard (where I still am). I love the center: It's really grounded in the libertarian perspective. Working with one of my colleagues, I organized a meeting between the directors of the Berkman Center and the directors of the West Point Army Cyber Institute to talk about cyber peace. It must have been like 2013 or something. And it was this fascinating moment where it was evident that both parties at the table actually shared a lot of common ground. We're talking about the same thing, but with such radically different languages and concepts, and radically different perspectives.

And I think that is what I'm aiming for with this idea of cyber peace, which is, if you're going to talk about stability, that's fine, you can call it stability. But the normal parties that you would convene when you talk about rules of the road in peacetime have to be at the table for the debate to be meaningful. You have to have a consideration for the tension between cyber power and human rights in peacetime. You have to have corporations at the table. What is the role of the private sector in relationship to the deployment of cyber power in peacetime?

All these other types of conversations are now starting to progress. We finally saw the private sector say, okay, maybe we do have a role in preserving peace and stability in peacetime. And we do have some form of responsibility in the face of cyber power. But that took a very long time.

CHRISTOPHER ANKERSEN: One of the questions I had written down was exactly that. If we look at the analogue, the world peace movement from the 60s, it shares a lot of the same ideas with the cyber utopian side. And civil society was a big driver there: NGOs, ordinary people, churches, and community groups, and there was a dialogue of sorts between the people and the government. It was reluctant, but

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

it worked in a way: The disarmament movement was a bottom up affair and it forced politicians to engage. But who wasn't involved in that conversation? Weapons manufacturers like Dow Chemical (the makers of napalm) and Raytheon. They were implicated in that conversation, but they were not really parties to it. They were like, "well, we'll wait and see, do we get an order next week? Or do we not but we don't really have a role in doing anything. We're not going to cut back if Ronald Reagan wants to engage more for the SDI then full speed ahead. Let someone else drive the ship. And we'll just provide what's needed." But this seems slightly different now that companies, corporations, and firms seem to, as you say, understand, at least implicitly, that they have more of a role.

But we don't see as much civil society involvement. People aren't on the streets out there looking for cyber peace. Do you think that that makes cyber peace a different kettle of fish and that we can't necessarily draw on past practices?

CAMILLE FRANÇOIS: There are so many interesting questions in what you just put on the table, I'll take at least three of them. The first one is: What is the private sector in this context? There isn't really one private sector. And when you think about it, you know, the Raytheon example is interesting, because you have the part of the private sector that is manufacturing and selling elements of cyber power. So the sort of "hacking for hire" types. And here, the debate is one of regulation. What is the appropriate regulation for shops that develop "zero-days for hire"? And that is a conversation that really was late to the party. We've seen organizations like the NSO Group go back and forth on what that means for them to meaningfully respect human rights. I think they got a lot of that very wrong. At the same time, though, regulators have been slow to catch up with that.

So there's a private sector in that way, that is part of this conversation, because it's one of regulation. Now, there is another private sector, which sometimes intersects, but mostly doesn't, which is the private sector on which this conflict is being deployed.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

And that raises a question of the role of a company like Microsoft, like Cloudflare, like Google, like Facebook. And here, what's really interesting is I have seen them be part of this conversation without acknowledging it, and therefore, we're missing the strategic guidance for it.

I'll give you a very bizarre, specific example, which is one that's really close to my heart. Ten years ago, Google launched my favorite feature anywhere on the Internet, which is the state sponsored warning. Google decided that its threat Intel team had the visibility to see when private citizens were being targeted by state sponsored actors on their services. And Google decided that it was worth telling these users and started rolling out a little message, initially in Gmail, that told its users "Google has reasons to believe that your account is being targeted by state sponsored actors." I spent a lot of time working on those features, and they are now replicated across the industry. Twitter's doing it, Facebook's doing it, and Microsoft's doing it. They're all saying not exactly the same thing and they're not all advising the same thing. But that is a hell of a recognition that in peacetime cyber power is deployed against the individual, and that there is a need to protect them and inform them.

CHRISTOPHER ANKERSEN: That is a great feature, but I would say most people don't know about it. Let's be honest, out of 7 billion people, probably less than 100,000 get that message, right? Because they're actually important enough in somebody else's ecosystem. And there are a few experts, such as yourself, who know about it, but that's what I mean. That's not the same as a peace symbol on a placard that a whole range of people might be attracted to and understand enough to, say, donate money to Greenpeace or actually go out and protest. It just seems to me that, in some sense, this is not a mass movement yet. There's a perfect example of technical capability to do it and some recognition among some people that it's necessary and possible. But does that include the people in the United States? Will Google warn somebody if they think the NSA or the FBI or someone is doing that? So few people know about that.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

It's not like, "Hey, man, like, you know, did you get your warning yet? Are you on the warning list?"

CAMILLE FRANÇOIS: I've worked with the targeted communities and the users who get the warning, and talked them through it. What do you understand about it? What did it feel like? What are your questions?

The targeted communities, they're exactly who you would expect: Members of parliament, elected officials, journalists, activists. I remember I did a user interview with a journalist in cybersecurity who eventually got the warning, and he said, "I finally got it! It was my badge of honor. I was the last one of my friends to get it. Now I can brag at DEF CON!" So there are communities for whom this is a known entity. But then I also talked to users who were more unaware: "Oh, yeah, I see this stuff. I think it's just routine stuff that they send to everybody, to keep people on their toes." They fundamentally don't understand this is because of exactly what you're saying, which is that we don't have a movement to explain it. What does it mean? What does it look like? What are the moments to panic and the moments to stay calm? And the advocacy piece, the civil society piece of it, has been quite slow to develop.

CHRISTOPHER ANKERSEN: You were going to talk about a third piece of the private sector before I interrupted?

CAMILLE FRANÇOIS: I was going to talk about the third piece of your question about the private sector, which is civil society. Last year, I joined the board of Digital Peace Now Society; I'm super excited about what they do. Their mission is to build up advocacy. But to be honest, I think that the fact that the research has been lagging behind has also hampered the advocacy movement's ability to develop. And I think that what's happened with Solar Winds is a good example. If you look at the cyber conversation, what do you see? People yelling at one another because they can't define what constitutes an attack. Which is okay, I understand. But it's really interesting because you can see that despite years of work on cyber conflict, those important terminologies about what can be expected and what isn't an appropriate response are still in flux, and they

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

remain contentious points in the actual academic literature. I think that this is because the academic focus on cyber peace for so long has been lagging behind the focus on cyber war.

CHRISTOPHER ANKERSEN: Do you think that part of this lag is not just on the research side, but because people perceive this to be "ones and zeros" and hacking and geeky and green screens and just weird stuff that they don't think they understand? Whereas, let's be honest, nobody understood nuclear weapons either, but they understood them enough to know "it goes boom, kills people: got it." And that was enough for people to get informed and have this grassroots "we don't want it anymore" type movement. Whereas with cyber there's some feeling of "Well, we need it; I don't really understand it; somebody knows better than me, the experts must have a hold on this." And so, therefore, even the civil society groups tend to be more informed, like EFF. These groups are a subset of the "geek community" that get it and therefore have concerns.

CAMILLE FRANÇOIS: It is a really interesting example. And lobbyists have been working with them for a long time. That's a conversation I've been having with them for ten years. EFF always says that that part (cyber peace) of the overall question isn't in their scope. So if you look, for instance, at the EFF statement on the Tallinn manual – it doesn't exist. That's not part of their scope.

So it's interesting to see that we can have entire conversations on norms that are applied to state power both above and beyond the threshold of armed conflict without any meaningful consultation of civil society organizations. Even EFF, which, as you said, is super tech savvy, isn't around the table. As a result, Tallinn 2, which is preoccupied by conflict below the threshold of armed conflict, has a chapter on human rights that is significantly smaller than the chapter on the Law of the Sea! The way we've been engaging with these questions, the way we've been defining the scope of these questions, is backwards.

CHRISTOPHER ANKERSEN: I wonder if that's because it comes from this idea, as you say, that most of the movement has come from the cyber security perspective, as opposed to the cyber

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

peace or cyber utopia side. Therefore, they see this as about securing stuff, protecting stuff, as opposed to liberating and kind of offbeat, as defining what we're actually trying to do, which is have a place where we can get stuff done.

CAMILLE FRANÇOIS: Exactly. What you are describing is a very tech-centric definition of cyberspace, one of tech bits and systems, which is why you care most about things like encryption. That very tech centric definition of the space has long been a problem for our ability to address wider issues such as peace and stability. That is, the problem that we had in 2016, in the face of Russian Foreign interference: both Silicon Valley and Washington were so preoccupied *excluding* that piece from their definition of cyber security. Again, from a normative perspective, perhaps that is okay, but at the end of the day, concretely, it means that in Silicon Valley, you had entire cybersecurity threat intelligence teams with not a single person in charge of detecting the attack that was going to come their way. So yes, you can have whatever definitions you want from a normative perspective. But this trickles down into how peace and security are actually cared for, and how we do defensive work in a way that leaves blind spots open and is, ultimately, problematic for peace and security.

CHRISTOPHER ANKERSEN: That is fascinating because it's this self-defined issue. Privacy? People get that and the solution to that is, somehow, more tech. Get a password manager, get a VPN, don't do this, don't do that. And platforms like Facebook will have a "real world harm threshold," which is to say that if somebody says they're going to murder somebody, we'll take that as a threshold to actually do something about it. But beyond that, on things like false information actually going to sway something, perhaps there has been too much of a free hand given, allowing companies to self-define, and therefore, opt out of these conversations. So it's not just that they're not welcome at the table, but they're also not necessarily knocking on the door to get to the table, either. They can sit back and say "we got this little gap here fixed and we got this little gap here."

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

But what about all "the rest of it"? And I think what you're saying is "all the rest of it" is cyber peace.

CAMILLE FRANÇOIS: Yes, it's not just hackers and "ones and zeros" everywhere. It's the unsexy but fundamental space where basic regulatory frameworks apply to protect peace and stability, how to define what's acceptable, what's not acceptable, who is in charge of defending it, and how we structure ourselves for it. What is the role of the private sector in that? What is the role of civil society? And what do we expect from our governments? Yes, it's not very sexy; it's not the hacker wars, but it represents the space where the vast majority of these incidents happen.

Because we're lacking this perspective, we're constantly getting blindsided by major events that after each of them, everybody says, "oh, how is it that we were possibly blindsided in this way?" My answer is that it's because our focus has been overly concerned with defining cyber war, the topic of countless doctrines, countless papers, and not focused enough on defining and organizing cyber peace.

CHRISTOPHER ANKERSEN: A last question then: What do you think the biggest threats are to this idea of cyber peace? Where would you say we were looking at the biggest barriers to actually getting to an idea of cyber peace?

CAMILLE FRANÇOIS: It's over indexing on offensive measures. It's that every incident that is getting in the way of peace and stability must be addressed by offensive measures, because our state of mind is that of cyber warfare and not that of cyber peace. Once you have a hammer, you have a hammer problem? What we need is a more positive, more defensive, broader understanding of cyber peace, across all of society. This last point is interesting because every time we confront a massive incident that was totally predictable, but yet not exactly in line with how we organize ourselves, one of the answers is, "oh, we need a whole of society response." That is true, but let's talk about why we don't have whole-of-society responses on things that touch cyber power.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

11

Overcoming Barriers to Empirical Cyber Research

Anne E. Boustead and Scott J. Shackelford

1 INTRODUCTION

Empirical studies have the potential to both inform and transform cyber peace research. Empirical research can shed light on opaque phenomena, summarize and synthesize diverse stakeholder perspectives, and allow causal inferences about the impact of policymaking efforts. However, researchers embarking on empirical projects in the area of cyber peace generally, and cybersecurity specifically, face significant challenges – particularly related to data collection. In this chapter, we identify some of the key impediments to empirical cyber research, and suggest how researchers and other interested stakeholders can overcome these barriers. While these issues stretch across different categories of research designs, some barriers are likely to generate more concern in the contexts of certain types of research questions, as is summarized in Table 11.1. Furthermore, while these obstacles are by no means unique to empirical cyber research, they are particularly salient in this context – and we focus on mechanisms for addressing these barriers that are most likely to be useful to cyber researchers.

2 BARRIERS TO EMPIRICAL CYBER RESEARCH

2.1 *Cyber Decisions and Outcomes Are Difficult to Observe*

Difficult-to-obtain data are a common and persistent problem for empirical cyber researchers. Although there are some publicly available data on cyber policies and outcomes (Federal Bureau of Investigation, 2020; Indiana Attorney General, 2020; National Conference of State Legislatures, 2020), these datasets can be fragmentary, and are few and far between. Data that have become available through less traditional means – such as the leaking of information after a data breach – can provide crucial insights into important, previously unobservable phenomenon, but their use in research raises novel and difficult ethical questions (Boustead & Herr, 2020). In the absence of publicly available datasets, researchers conducting empirical cyber

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

TABLE 11.1 *Most salient barriers to addressing different types of empirical cyber research questions*

Type	Description of Type	Example Cyber Question	Most Salient Barrier
Exploratory	Focus is on describing and explaining phenomena; may be used to analyze the range of variation in a phenomenon	How do organizations decide whether to use an external cyber risk decision-making framework?	*Empirical cyber research projects frequently require expertise from multiple domains*, complicating systematic exploration of cyber phenomena
Parameter Estimation	Focus is on quantitatively estimating characteristics of a population in a statistically valid way; generally requires particular kinds of random sampling	How many hours of cybersecurity training do hospital employees receive every year?	*Research may only be possible in a narrow range of contexts*, making it difficult to systematically observe a population of interest
Causal	Focus is on establishing whether a cause-and-effect relationship exists between two characteristics of a phenomenon	Do policies requiring regular password changes reduce the frequency of successful cyberattacks?	*Cyber decisions and outcomes are difficult to observe*, making it difficult to identify and evaluate policymaking

projects must rely more heavily on data collection, increasing the time, effort, and resources necessary to conduct research.

Data collection in empirical cyber research is further complicated by the range of actors involved in cyber policy, and differences in how these actors document and disclose their cyber decision-making. Government cyber policymaking is typically memorialized in publicly released documents – including statutes and judicial opinions – which can be analyzed and used to evaluate the effects of these policies on important outcomes (Romanosky, Telang, & Acquisti, 2011). However, much cyber policymaking occurs on an organizational level through decisions made by specific companies and groups about how to manage their own cyber practices (Harknett & Stever, 2009). This decision-making frequently does not result in public documentation, and organizations may be highly reticent to disclose details of their cyber practices due to concerns about security, brand reputation, or liability.

Researchers cannot evaluate policies that they cannot observe and, perhaps more insidiously, efforts to evaluate observable government policies may be undermined by simultaneous and unobservable organizational decision-making. For example, a heavily publicized data breach event could result in observable legislation mandating

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

employee cybersecurity training, as well as unobservable changes in corporate cyber infrastructure. If the frequency of data breaches declines after the legislation becomes effective, researchers may attribute this change to the legislation without being aware of the confounding and unobservable changes in corporate cyber infrastructure. Reluctance to provide information about cyber decision-making can also result in low survey response rates, making it difficult to accurately estimate how often organizations are adopting particular cyber practices.

2.2 *Empirical Cyber Research Projects Frequently Require Expertise from Multiple Domains*

Cyber systems consist of more than just technology; they also include the people and organizations involved in using and managing cyber systems. Consequently, empirical cyber research often requires data and analytic techniques from multiple domains and disciplines. For example, a project studying how the passage of data breach notification laws impacts cybersecurity behaviors and outcomes would require expertise in law, behavioral sciences, and computer science (Murciano-Goroff, 2019). The range of expertise necessary to conduct these projects generally suggests the need for an interdisciplinary research team. However, differences in the expectations and incentives placed upon researchers in different disciplines may make collaboration difficult.

2.3 *Research May Only Be Possible in a Narrow Range of Contexts*

While some categories of research questions can be answered with only a limited range of observations, others require either a broader scope of data collection or the use of specialized sampling techniques. This is particularly important when trying to describe a characteristic of a population; for example, when estimating the percentage of Fortune 500 companies that employ a Chief Information Security Officer, or how many hours of cybersecurity training hospital employees receive every year. In order to estimate these characteristics in a statistically valid way, researchers must be able to select individuals from the population to observe so that (1) every member of the population could potentially be studied, and (2) the researcher knows how likely it is that each member would be selected. This process – which is known as conducting a probability sample – generally requires identifying every member of the population and selecting members at random to observe (Groves et al., 2011). In the case of cyber peace research, identifying every member of the population can be particularly difficult, especially when researchers are trying to estimate characteristics of technical populations (such as malware) rather than human ones. Even when it is possible to address a research question by studying a narrower population, this choice may impact the generalizability of the research (Lee & Baskerville, 2003). As a result, both researchers and policymakers must be careful when trying to

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

generalize the results of the study. For example, further research would be needed to determine whether the results of a survey of cybersecurity practices conducted in Indiana could be generalized to other states (Boustead & Shackelford, 2020).

3 OVERCOMING BARRIERS

Although these barriers pose significant challenges to empirical cyber research, they are not insuperable. In the remainder of this document, we identify several practices that individual researchers, universities, and other organizations could adopt to facilitate empirical cyber research.

3.1 *Incentivize Interdisciplinary Research Teams*

To overcome these difficulties, exploratory cyber research projects may especially benefit from an interdisciplinary team, with expertise in technology, policy, law, and behavioral science. Fortunately, there is a long history of interdisciplinary collaboration in cyber research, including cross-disciplinary conferences, journals, academic programs, and other initiatives. In order to further encourage interdisciplinary cyber research, we would suggest that academic leaders in multiple disciplines make clear how interdisciplinary research will be accounted for during the tenure and promotion process (Benson et al., 2016). Additionally, researchers across multiple disciplines should be encouraged to engage in cross-disciplinary teaching experiences in order to educate future researchers and decision-makers to engage in interdisciplinary research, and create partnerships between disciplines to facilitate future research. An example of this approach in action is the IU Cybersecurity Clinic, which is unique in both its interdisciplinary breadth, as well as the fact that it is open to all graduate students across campus and offers applied service-learning opportunities to assist local and state-level critical infrastructure providers.

3.2 *Partnerships Are Key*

Oftentimes, empirical cyber research questions may be of interest to a variety of stakeholders in the public and private sectors. A state government may be interested in information about the uptake of cybersecurity practices amongst businesses in their jurisdiction, while a trade group might be interested in perceptions of privacy protections amongst their constituents. For example, the authors of this paper have collaborated with the State of Indiana to field a survey on cybersecurity practices amongst organizations in Indiana in order to address both academic and policy questions on cybersecurity decision-making. Under these circumstances, partnering with stakeholders has the potential to facilitate and improve cyber research. Research partners can provide insights into the phenomena in which they are involved, and insider knowledge about how policies are implemented in practice

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

can provide a critical counterpoint to academic expertise. Furthermore, stakeholders are often experts in their own decision-making, and emic explanations about their policies and practices can be irreplaceable.

Research partnerships with public or private stakeholders can take on a number of forms. Researchers can consult with stakeholders during project development in order to identify potential causal mechanisms, locate existing data sources, and preview interview questions to determine whether they are likely to elicit relevant information. Stakeholder research partners may also be willing to facilitate data collection by distributing surveys or providing introductions to potential interview subjects. Because they are likely interested in the results of research, stakeholder partners may also be helpful in disseminating the results of research projects and encouraging consideration of policy recommendations resulting from the project.

While partnerships with public or private organizations can greatly benefit empirical cyber projects, researchers must be mindful of several potential complications. Public and private organizations may have a more limited remit than the population that might be of interest to the researcher. For example, a state or local government may be able to provide data about their own jurisdiction, and an industry trade group may be able to assist in distributing a survey to their members. These constraints can generally be addressed by narrowing the research question to focus on the population for which data are available; however, a more limited study may be less generalizable, and efforts to use these studies for policy decision-making in other areas must account for differences in context. It may also be helpful to repeat research in multiple contexts in order to explore the circumstances under which the results of the study hold.

Researchers who partner with public or private entities should also be prepared to navigate potential conflicts between the goals of the research partner and the goals of academic research. Organizations may partner with researchers because they have an interest in obtaining answers to particular questions or learning more about phenomena that affect them. Researchers may consider expanding the scope of their research to ensure that questions of interest to the partner are also addressed, and seeking out partnerships where there is a natural overlap in the questions of interest. However, partners should never have control or veto power over whether the results of the research are released. In order to ensure that partnering organizations can benefit and learn from the research, researchers should consider ensuring that results are available in formats that are usable by the partner; for example, publishing reports and podcasts, as well as journal articles.

3.3 *Publish Cyber-Related Data*

The field of empirical cyber research as a whole would benefit tremendously from an increase in the scope of publicly available data on cyber policies, decisions, and outcomes. Publicly available data facilitate and incentivize research by lowering the

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

costs of undertaking projects. They also create efficiencies by ensuring that data collected are available to many researchers, reduce the burden on participants who may be asked to participate in multiple studies unless data collection is coordinated, and increase transparency in both research and policymaking (Napoli & Karaganis, 2010). Organizing the release of datasets could also serve as a mechanism for promoting high-quality cyber research if the data released are valid and reliable.

There are a number of mechanisms for ensuring the availability of empirical data on cyber phenomena. Over the short term, the publication of an annotated bibliography describing the datasets that are available, and highlighting where the collection of data in other domains has touched upon cyber-related issues, would both make those data more accessible to researchers and identify gaps in current data availability. Efforts could then be undertaken to expand current data collection projects to include information about cyber-related issues where relevant; for example, adding a question to a survey of hospital administrators to ask about their cybersecurity practices. Finally, surveys and other data collection projects focused on cyber issues could be undertaken and expanded, with priority given to efforts that can be repeated on a yearly basis in order to observe changes over time. These efforts could be facilitated through collaboration with existing public–private cyber partnerships, such as Executive Councils on Cybersecurity and organizations designed to share cyber threat information within sectors, such as information sharing and analysis centers and information sharing and analysis organizations. There is no one-size-fits-all model, but through experimentation and deeper partnerships, we may glean a more accurate picture of the cyber peace landscape.

REFERENCES

Benson, M. H., Lippitt, C. D., Morrison, R., Cosens, B., Boll, J., Chaffin, B. C., … Link, T. E. (2016). Five ways to support interdisciplinary work before tenure. *Journal of Environmental Studies and Sciences*, 6(2), 260–267.

Boustead, A. E., & Herr, T. (2020). Analyzing the ethical implications of research using leaked data. *PS: Political Science & Politics*, 53(3), 505–509.

Boustead, A. E., & Shackelford, S. (2020). State of Hoosier Cybersecurity. Retrieved from www.ibrc.indiana.edu/studies/State-of-Hoosier-Cybersecurity-2020.pdf?_ga=2.164292790.820309617.1609805269-789814884.1603708149

Federal Bureau of Investigation. (2020). 2019 Internet Crime Report. Retrieved from https://pdf.ic3.gov/2019_IC3Report.pdf

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology* (Vol. 561). John Wiley & Sons.

Harknett, R. J., & Stever, J. A. (2009). The cybersecurity triad: Government, private sector partners, and the engaged cybersecurity citizen. *Journal of Homeland Security and Emergency Management*, 6(1), 1–14.

Indiana Attorney General. (2020). Identity Theft Prevention. Retrieved from www.in.gov/attorneygeneral/2874.htm

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, 14(3), 221–243.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Murciano-Goroff, R. (2019). Do Data Breach Disclosure Laws Increase Firms' Investment in Securing Their Digital Infrastructure? Paper presented at the Workshop on the Economics of Information Security.

Napoli, P. M., & Karaganis, J. (2010). On making public policy with publicly available data: The case of US communications policymaking. *Government Information Quarterly*, 27(4), 384–391.

National Conference of State Legislatures. (2020). Security Breach Notification Laws. Retrieved from www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

Romanosky, S., Telang, R., & Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2), 256–286.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

12

Bits and “Peaces”

Solving the Jigsaw to Secure Cyberspace

Stéphane Duguin, Rebekah Lewis, Francesca Bosco, and Juliana Crema[*]

1 INTRODUCTION

Efforts to create peace in cyberspace can, at times, be much like trying to assemble a 1,000 piece jigsaw puzzle without a picture of the finished product, full of important, related elements, but lacking an overall strategy. Much like the missing picture on the puzzle box, the absence of a mutually agreed-upon definition of “cyber peace” is itself one of the fundamental challenges to achieving it. Without a common understanding – a common vision – it is difficult to come together and work collectively toward a common goal. While agreeing on a universal definition of any truly global concept is inherently challenging (witness the ongoing debates surrounding “sustainability”), due to the sheer number and diversity of perspectives involved establishing a shared understanding of cyber peace is particularly difficult due to the complex and evolving nature of cyberspace, and the nature and meaning of peace itself. By taking a more operational perspective in this chapter and building from the work of others throughout this edited volume, our hope is to advance the discussion on cyber peace beyond this uncertainty – in essence, to transcend it.

First, we will set forth a “light-weight” operational definition of cyber peace that we believe is compatible with more theoretical formulations of the concept, while providing a guiding compass point for both strategic and tactical activities. To be impactful, we argue that any approach to cyber peace must, above all, be concerned with human well-being and, therefore, contemplate the integrated, multidimensional components of the human experience. As outlined in the first chapter of this volume, there are various interpretations of cyber peace. Some understand it solely as a concept to be theorised, whereas others consider it to be a set of practices

[*] The CyberPeace Institute is an independent, nonprofit organization headquartered in Geneva that works to enhance the stability of cyberspace by decreasing the frequency, impact, and scale of destructive cyberattacks against civilians. The Institute promotes transparency about such attacks and holds malicious actors to account for the harm they cause on vulnerable communities and populations. To this end, the Institute’s mission is to ensure a human-centric and evidence-led response to cyber operations.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

that can be employed (Marlin-Bennett, 2022, pp. 4–6). With this work in mind, we will build upon Marlin-Bennett's conception of cyber peace as a practice in order to better understand the human role and the related impact we can have to promote cyber peace and accountability in cyberspace. Second, we will highlight two key challenges that we believe must be overcome on the road to cyber peace. In assessing these challenges, we also seek to bring to the fore a broader geopolitical issue, the growing and fundamental redistribution of power that is not supported by a complementary redistribution of oversight and accountability. Lastly, we will argue that the principle of accountability – as a generally applicable concept and a key component of literature on institutional analysis such as the Ostrom Design Principles – provides a flexible and durable means to pursue cyber peace. By taking into account this operational understanding of cyber peace and by using current examples to illustrate how they apply can help to further guide the path toward a sustainable cyber peace framework.

2 DEFINING CYBER PEACE

In an effort to highlight the necessary collective approach to achieve cyber peace, we have built our operational understanding of cyber peace around previous work, but have adjusted the focus to ensure that it is human-centric. As discussed in the Preface and first chapter of this edited volume, discussions of cyberspace, operations, security, and peace have come from a variety of actors in a move toward a multistakeholder approach to cyberspace and away from the previous focus upon state-centered security models (Shackelford, 2022, p. xxv). The state as a focal point makes sense in a Westphalian world order in which national territory and governments serve as the primary mechanisms for protection of rights and preservation of stability and order. But the scope of cyberspace – as a "notional environment" defined by connected networks and devices – is rapidly expanding (Delerue, 2020, p. 29). Today, from a micro-level perspective, computers, networks, and, information and communication technologies – "cyber" – is woven into almost every aspect of human life. Equally important to recognize is the macro-level perspective in order to understand how the complexities of our digital life are nestled into broader societal and geopolitical contexts. Accordingly, we assert that any efforts to achieve cyber peace should and must, as a moral imperative, be centered around and motivated by a concern for the well-being of *individual human beings* in order to achieve a peace beyond the mere absence of war (Diehl, 2019, pp. 2–3). Echoing Heather Roff (2016, p. 8), we believe that the individual human should be the main referent for a guiding conception of cyber peace. In keeping with this singular focus on the human being as the center point of a peaceful cyberspace, we propose that *cyber peace exists when human security, dignity, and equity are ensured in digital ecosystems.*

This formulation of cyber peace is intended to be highly actionable at an operational level and, we believe, is complementary to and compatible with existing

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

related scholarship which has been previously discussed in this volume.[1] For those seeking to actively pursue cyber peace, the definition is intended to be instructive on a number of practical levels and works to approach these issues from a human perspective from the start. In this way, we can begin to address the challenges and obstacles in achieving accountability in cyberspace. In an effort to clarify each element of cyber peace listed above, there are specific criteria and questions to consider. For example, we believe that human security exists in cyberspace when services essential to human life and related critical infrastructure are protected. Based on this definition, we can begin to think through cyber-related topics by using a human-centric lens, and by questioning which rights and freedoms have been violated such as, "… the right to life, liberty and security of person" (United Nations, 2015, p. 8). Some questions to think about include whether there has been an obstruction to essential resources and services; this line of critical thinking will also begin to highlight the question of accountability, and in cases of attacks against healthcare facilities; for example, who holds responsibility for the failure to protect the element of human security.

Moreover, following the foundation laid out by human security, in the cyberspace context human dignity presents a mutually reinforcing concept as these associated rights rest upon the fulfillment of security in cyberspace. With this in mind, human dignity exists in cyberspace when individual's beliefs, cultural rights, and ability to participate in society are protected. Human dignity is unique to the individual's experience and context-specific to their everyday realities. Rights relating to this definition include, but are not limited to, civil and political rights, along with freedom of expression and assembly, as well as cultural and indigenous rights. Furthermore, human equity exists in cyberspace when individuals are protected against discrimination, bias, prejudice, and inequality. The importance of human equity in cyberspace stems from the reality that not everyone is starting at the same position in life and that these discrepancies need to be rectified in order for cyber peace to exist. This understanding follows the first and key tenant from the Universal Declaration of Human Rights, which emphasizes that "all human beings are born free and equal in dignity and rights" (United Nations, 2015, p. 4). This definition helps to get to the root problems that need to be addressed in order to resolve inequalities and can include issues such as political or developmental barriers to equity, or social constructions which inhibit upon one's rights. This holistic and proactive approach is needed to ensure that these barriers are eliminated for people and communities everywhere. We view cyberpeace as encompassing three distinct elements: human security, dignity, and equity. These key elements relate to various dimensions of the human experience, including political, economic, and social considerations,

[1] For example, this definition comports with the four pillars of a positive cyber peace "…as a system that: (1) respects human rights and freedoms; (2) spreads Internet access along with cybersecurity best practices, (3) strengthens governance mechanisms by fostering multi-stakeholder collaboration, and (4) promotes stability and relatedly sustainable development" (Shackelford, 2020, pp. 15–16).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

and in this way are closely linked with human rights. To be clear, these three key elements are intertwined, interdependent, and intersectional as a necessary effort to achieve cyber peace. The human rights specifically encompassed by human security, dignity, and equity build upon and reinforce each other, as is relevant for each individual's experience.

Keeping these concepts in mind, but further building upon our understanding of cyber peace, the role of accountability becomes much more apparent. By grounding these definitions in rights and freedoms, and while maintaining a human-centric perspective, we can further question the intersection of the virtual and physical worlds, and the role that each actor plays in these ecosystems. Having a clearer understanding of the roles and responsibilities, both on and offline, will help to rectify the accountability deficit we currently face due to the rapid evolution and convergence of disruptions in technology, geopolitics, and human behavior.

One example of the operationalization of our approach toward cyber peace is the focus we have been devoting to the healthcare sector. As the extent of people relying on health services for necessary human needs increases, the potential harm to human security and dignity are immense. Malicious cyber operations against healthcare facilities put human lives in jeopardy and require immediate action. To this end, we supported a call on the world's governments to collaboratively work to stop cyberattacks against healthcare facilities and related critical infrastructure entities. Then, considering the increasing gap reported between the variation and sophistication of cyberattacks and the ability for healthcare sector entities to protect themselves from such attacks, we set up Cyber 4 Healthcare. This initiative is a global matchmaking service to partner civil society organizations and healthcare providers with private sector actors to individually assist them in protecting their services in order to decrease their vulnerability to cyberattacks, while considering their local context. The personalized advice and discussions through Cyber 4 Healthcare is just one example of how cyber peace, as it encompasses human security, dignity, and equity, must truly span the globe, inside and out, while maintaining contextual relevance.

3 KEY CHALLENGES ON THE ROAD TO CYBER PEACE

In order to further unpack the goal of cyber peace through accountability, we must be cognizant of the challenges and obstacles in this realm. In order to illustrate this, we have identified two deeply rooted and largely false assumptions about the nature of cyberspace itself that must be debunked and counteracted in order to make meaningful progress toward cyber peace. First, we must recognize the unequal and disproportionate access and engagement with cyberspace around the world, and address this issue in the discourse around responsibilities and responsible behavior in cyberspace. Second, we must acknowledge and tackle head-on, through creative, out-of-the-box thinking, the persistent tensions and gaps in the existing ecosystem of laws, norms, and principles governing cyberspace, and the use of information and

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

communications technologies (ICTs). By analyzing these issues through a cyber peace lens, we can begin to address them on the basis of rights and freedoms afforded to all in international treaties and declarations.

3.1 *Access and Security in Cyberspace*

Keeping in mind the definition of cyber peace and its corresponding elements, access and security in cyberspace remains a prominent challenge. Communities around the world are in vastly different stages of development and implementation when it comes to cyberspace infrastructure and technology, which thus leads to questions of human equity and the impact that this discrepancy in development has upon end-users and citizens of the world more generally. Bias and subjectivity are hard-wired problems in technology, though access to this technology is in itself unevenly distributed, which deepens existing inequalities. In order to keep this issue in a global context, it is also highlighted by the UN's Sustainable Development Goals, particularly goal number 9, which is to "build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation" (United Nations Department of Economic and Social Affairs, 2020, p. 42). Specifically in relation to access to the Internet, the UN cites that, "in 2019, almost the entire world population (97%) lived within reach of a mobile cellular signal, and 93% lived within reach of a mobile-broadband signal" (United Nations Department of Economic and Social Affairs, 2020, p. 43). However, despite this high percentage of coverage the UN found that, "most of the offline population live in LDCs, where only 19% use the Internet …" (United Nations Department of Economic and Social Affairs, 2020, p. 43). Moreover, the 2019 Human Development Report emphasizes this point and warns that, "… while access to basic technologies is converging, there is a growing divergence in the use of advanced ones …" leading to a growing concern about a so-called New Great Divergence, following the first divergence created by the Industrial Revolution (United Nations Development Programme, 2019, pp. 200–203). With a greater emphasis placed upon the question of human equity and an active approach to facilitate the participation of all in cyberspace, concerns about the digital divide can begin to be addressed. As stated in the Sustainable Development Goals, bridging the digital divide by providing Internet access to the 3.6 billion people – nearly half of the world's population – in the developing world who are not online "is crucial to ensure equal access to information and knowledge, as well as foster innovation and entrepreneurship" (United Nations Development Programme, n.d., para. 3). Moreover, disparate levels of exposure and access to technology mean communities have vastly different experiences upon which to form mature policy positions, significantly affecting their ability to participate meaningfully in global fora, and therefore harming their overall security as a citizen and as a person.

To be clear, getting online is only one piece of the puzzle. While infrastructure is a first and crucial step to access, it is not enough simply to invest in the installation

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

of fiber or cell towers, or even to foster an ecosystem of service providers. Once online, users must be able to engage and act without threat to their privacy, freedom of speech, and financial or physical security. Such threats, in the form of online discrimination, censorship, manipulation, and surveillance are faced by vulnerable populations around the world, but manifest differently depending on the relevant technology and context specific to them. In order to be sustainable and effective, a cyber peace framework must, therefore, acknowledge and account for the distinct ways that cyberattacks impact different populations depending on their context and unique situations, and therefore threaten their human security. The role of accountability becomes clearer in this context, because by recognizing threats to an individual's opinion or their privacy, the behavior of states, industry, civil society, and end users becomes more apparent.

The healthcare sector in the context of the COVID-19 pandemic presents one such case where vulnerable communities, whose ability to access and securely engage in cyberspace, have been severely compromised as a result of specific circumstances and characteristics. Long before the coronavirus outbreak, the healthcare sector's dependence on digital technology and connectivity had skyrocketed. This dependence, combined with the sensitive data and services under its purview, put the healthcare sector at high-risk to cyberattacks, such as ransomware or data breaches. Following the declaration of a global pandemic and the sudden increase in demand for medical facilities and services, this community became even more vulnerable to existing security threats as they scrambled to set up field hospitals and testing centers, produce and procure equipment, and reshuffle staff and schedules. A well-publicized attack against a hospital in Düsseldorf, Germany, forced the hospital to turn away patients, including a woman who later died, due to a ransomware attack that encrypted thirty of the hospital's servers (Goodin, 2020). In cases like this, by disrupting healthcare operations, such attacks have a very real and tangible impact on the health and well-being of its staff, patients, and the broader community the healthcare sector serves. It shows how questions and concerns over human security should be at the forefront of cyber peace since, at the end of the day, events such as cyberattacks against hospitals have an impact on human lives and their overall well-being.

In addition, the COVID-19 infodemic is another closely related example of the unique impact of cyberattacks on specific communities. These communities are often not defined by any geographic or territorial boundaries, but are still protected under the concepts of human security, dignity, and equity. Due to the nature of the COVID-19 outbreak:

> … communities are relying on online resources to be informed, and are producing information on their own. This leads to a massive generation of online content, blending information coming from official channels (media outlets, international organization bodies, governments), private communication entities and user's generated content (CyberPeace Institute, n.d., para 1).

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

The World Health Organization (WHO) has identified this "blending of information" as an "infodemic," defined as, "… an over-abundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it" (World Health Organization, 2020, p. 2). For example, as an increasing number of people turn to online resources to work and study from home, malicious actors are taking advantage of this influx of online activity. In one case, the WHO itself was hacked and phishing emails that mimicked the organization's internal email system were sent out by a malicious actor (Satter et al., 2020).

These kinds of attacks not only weaken public trust in authoritative institutions like the WHO, but also cause these organizations to divert staff resources away from their usual activities to respond to attacks and mitigate their effects.[2] Beyond the community of institutions like the WHO, this infodemic greatly impacts the broader community of so-called "netizens" – engaged and responsible online users – by eroding their sense of trust and security on the Internet itself. Without a sense of security online, those who are already vulnerable to attacks or influence are left more vulnerable, and any sense of accountability is lost. These are just two specific examples of how global events and changing circumstances, even those – like the COVID-19 pandemic – with no direct relationship to digital technology, can quickly create new vulnerabilities and threats to online access and security. In pursuing cyber peace, we must account for this volatility and incorporate mechanisms to protect vulnerable populations as they arise in a rapidly changing global landscape.

4 THE ECOSYSTEM OF LAWS AND NORMS

The current ecosystem of international law and norms surrounding cybersecurity is complex to say the least. While these complexities present many areas of interesting debate, specifically about polycentric engagement, those in pursuit of cyber peace must work to identify and address the gaps and ambiguities that have the greatest impact on civilian life and human well-being with relation to human security, dignity, and equity. The COVID-19 pandemic again provides a powerful recent example of some of the impact of such gaps and ambiguities.

Cyberattacks against hospitals, such as the one in Düsseldorf as previously discussed, and other facilities during the pandemic emphasize the importance of protecting essential services – especially (but not only) during times of crisis when the civilian population is particularly dependent upon them and their security is at risk. However, both international law and norms present hurdles. Related to international law, the question of attribution presents a foundational issue regarding the ability to track adherence to specific responsibilities and bring claims against specific states. In

[2] See the following public service announcement as an example: www.who.int/about/communications/cyber-security.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

addition, ongoing debate regarding relevant thresholds for violations of obligations related to territorial sovereignty and due diligence also frustrates the ability to bring substantiated claims (Open Ended Working Group, 2020, p. 5). Voluntary nonbinding norms that have been proposed to support or complement existing legal obligations are also challenged by ambiguity regarding the meaning of certain key terms, including critical civilian infrastructure – as evidenced by debates and comments at the Open Ended Working Group (OEWG) and discussions regarding a new norm prohibiting attacks against medical facilities further underline this ambiguity (International Committee of the Red Cross, 2020).

This latter issue regarding critical infrastructure is highlighted by the norms outlined in the 2015 report by the UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (UN GGE) (United Nations General Assembly, 2015). Two of these norms address the protection of "critical infrastructure," which is of particular importance to discussion and analysis of the implications of cyberattacks against the healthcare sector, and specifically in how they relate to human security, dignity, and equity:

> (f) A State should not conduct or knowingly support ICT activity contrary to its obligations under international law that intentionally damages critical infrastructure or otherwise impairs the use and operation of critical infrastructure to provide services to the public.
> (g) States should take appropriate measures to protect their critical infrastructure from ICT threats, taking into account General Assembly resolution 58/199 on the creation of a global culture of cybersecurity and the protection of critical information infrastructures, and other relevant resolutions … (United Nations General Assembly, 2015, p. 8).

Not only is the definition of critical infrastructure itself the subject of much debate, the question of what constitutes "appropriate measures" in the context of norm (g) is also unclear. This snapshot of how the existing legal and normative framework for cyberspace applies to a specific sector in a time of acute crisis demonstrates some of the current gaps between conceptualization and on-the-ground reality. Without clarity regarding these foundational components, the effectiveness of law and norms as mechanisms for change and accountability will be limited. In the meantime, the human cost and impact of these attacks continue to rise as individual's security, dignity, and equity is threatened as the pandemic rages on.

5 ACHIEVING CYBER PEACE

In keeping with the notion of cyber peace as a multidimensional concept, adopting a general theory of change rather than attempting to enumerate specific measures will maximize operational flexibility, durability, innovation and, ultimately, impact. In critiquing the World Federation of Scientist's Erice Declaration, which

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

applies a top–down governance solution to cyber peace, Heather Roff notes that "by framing the issue this way, the Scientists discount problems associated with unjust social structures, as well as the unsatisfactory nature of the entire international legal framework" (2016, p. 5–6). This is but one example of how prescribing specific approaches – in this case, peace through legal governance, may discount important issues that specifically relate to human security, dignity, and equity. Another point of reference to consider are the principles put forth by Ostrom which show "… in many places around the world how communities devise ways to govern the commons to assure its survival for their needs and future generations" (Walljasper, 2011). These principles can be used to form sustainable and equitable governance systems in communities which form an integral part of the polycentric governance model discussed previously in this edited volume. In essence, the principles put forth by Ostrom are a way to assess one's responsibility to act in their community, so that future generations may also enjoy the same natural resources; for example (Walljasper, 2011).

As applied to cyberspace, we believe a general theory of accountability can provide the needed flexibility and durability to serve as a foundation for a globally applicable methodology for achieving cyber peace. Such a theory of *accountability* is not synonymous with *attribution* or so-called "naming and shaming." Rather, we recommend a very practical understanding of accountability as used in a variety of everyday settings, from the hyper-local to the international, building upon the polycentric approach discussed throughout this volume. For example, in the maintenance of a dwelling, training of a sports team, or employees at a coffee shop – in all these settings, specific actors have clear responsibilities and roles aimed at achieving a common goal and each are held accountable for these actions through various mechanisms. Accountability requires an evolving understanding of relevant stakeholders and responsibilities. More specifically, at the CyberPeace Institute, we believe that a systematic approach to accountability involves the following key steps for each stakeholder; identification of relevant responsibilities, confirmation of commitment to these responsibilities, tracking or measurement of adherence to these responsibilities, and analysis and implementation of effective measures to ensure or increase adherence. We believe that these four steps complement existing work on the topic of cyber peace by advocating for both a bottom–up approach to governance, as outlined in the polycentric model, but by also promoting a simultaneous top–down approach in governance to ensure that appropriate regulation and oversight works to promote accountability of all stakeholders in cyberspace.

6 CONCLUSION

The key challenges above expose an underlying redistribution of power as a result of changing digital ecosystems; a redistribution that is not accompanied by equally robust mechanisms for accountability that can be leveraged to protect individual human beings and their rights and freedoms, both on and offline. By defining cyber

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

peace around human security, dignity, and equity we can take direct aim at this systemic problem and begin to address the human impact of infringements upon these fundamental building blocks of peace.

As we move into a brave new world, we want to actively and deliberately design our future. Cyber peace is a way to articulate the desired contours of that future and provide clear compass points toward a destination that will benefit all. Recognizing again that common action requires common understanding and a common goal, we must be clear about what we are after and why. The CyberPeace Institute is committed to further operationalize the concept of cyber peace. Such operationalization does not require consensus regarding a finite list of the specific means to achieve our end goal. With the rough contours and a working theory of accountability, we can move forward in a common pursuit of cyber peace.

REFERENCES

CyberPeace Institute. (n.d.). What is the infodemic? Retrieved from: https://cyberpeaceinstitute.org/blog/2020-03-25-what-is-the-infodemic

Delerue, F. (2020). *Cyber operations and international law*. Cambridge University Press. DOI: 10.1017/9781108780605

Diehl, P. F. (2019). *Peace: A conceptual survey*. Oxford Research Encyclopedia of International Studies. Oxford University Press. https://doi.org/10.1093/acrefore/9780190846626.013.515

Goodin, D. (2020, September 19). A patient dies after a ransomware attack hits a hospital. *Wired*. Retrieved from: www.wired.com/story/a-patient-dies-after-a-ransomware-attack-hits-a-hospital/

International Committee of the Red Cross. (2020, February 11). Norms for responsible State behavior on cyber operations should build on international law. Retrieved from: www.icrc.org/en/document/norms-responsible-state-behavior-cyber-operations-should-build-international-law

Marlin-Bennett, R. (2022). Cyber Peace: Is that a thing? In S. J. Shackelford, F. Douzet & C. Ankersen (Eds.), *Cyber Peace: Charting a path toward a sustainable, stable, and secure cyberspace* (pp. 3–21). Cambridge University Press.

Open Ended Working Group on developments in the field of information and telecommunications in the context of international security. (2020, May 27). Second "Pre-draft" of the report of the OEWG on developments in the field of information and telecommunications in the context of international security. Retrieved from: https://front.un-arm.org/wp-content/uploads/2020/05/200527-oewg-ict-revised-pre-draft.pdf

Roff, H. M. (2016). Cyber Peace: Cybersecurity through the lens of positive peace. *New America*. Retrieved from: https://static.newamerica.org/attachments/12554-cyber-peace/FOR%20PRINTING-Cyber_Peace_Roff.2fbbb0b16b69482e8b6312937607ad66.pdf

Satter, R., Stubbs, J., & Bing, C. (2020, March 23). Exclusive: Elite hackers target WHO as coronavirus cyberattacks spike. *Reuters*. Retrieved from: www.reuters.com/article/us-health-coronavirus-who-hack-exclusive/exclusive-elite-hackers-target-who-as-coronavirus-cyberattacks-spike-idUSKBN21A3BN

Shackelford, S. J. (2020). Inside the global drive for Cyber Peace: Unpacking the implications for practitioners and policymakers. SSRN. Retrieved from: https://ssrn.com/abstract=3577161 or http://dx.doi.org/10.2139/ssrn.3577161

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Shackelford, S. J. (2022). Introduction. In S. J. Shackelford, F. Douzet, & C. Ankersen (Eds.), *Cyber Peace: Charting a path toward a sustainable, stable, and secure cyberspace* (pp. xix–xxxi). Cambridge University Press.

United Nations Department of Economic and Social Affairs. (2020). The Sustainable Development Goals Report 2020. Retrieved from: https://unstats.un.org/sdgs/report/2020/

United Nations Development Programme. (2019). Human Development Report 2019: Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century. Retrieved from: http://hdr.undp.org/sites/default/files/hdr2019.pdf

United Nations Development Programme. (n.d.). Goal 9: Industry, innovation and infrastructure. Retrieved from: www.undp.org/content/undp/en/home/sustainable-development-goals/goal-9-industry-innovation-and-infrastructure.html

United Nations, General Assembly. (2015, July 22). Report of the group of governmental experts on developments in the field of information and telecommunications in the context of international security, A/70/174. Retrieved from: https://undocs.org/A/70/174

Walljasper, J. (2011, October 2). Elinor Ostrom's 8 Principles for Managing a Commons. *On the Commons*. Retrieved from: www.onthecommons.org/magazine/elinor-ostroms-8-principles-managing-commmons

World Health Organization. (2020). Novel Coronavirus (2019-nCoV): Situation Report – 13. Retrieved from: www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

13

Cyber Hygiene Can Support Cyber Peace

Megan Stifel, Kayle Giroud, and Ryan Walsh[*]

Among high-profile cybersecurity incidents over the past decade, several were reportedly the work of nation-state actors. The actors leveraged tactics, techniques, and procedures to take advantage of known vulnerabilities – technical and human – to undertake actions that compromised personal information, risked human health, and paralyzed the global supply chain. Left unchecked, the scale and breadth of such actions can threaten international stability. Yet, an examination of high-level cases suggests that basic cyber hygiene is an accessible and practical approach to mitigate such incidents, enhance confidence in the use of information and communications technology (ICTs) and, ultimately, advance cyber peace.

[*] About the Global Cyber Alliance.

Founded in 2015 by the District Attorney for New York, the City of London Police, and the Center for Internet Security, the Global Cyber Alliance (GCA) is a charitable organization dedicated to reducing cyber risks. The GCA accomplishes this mission by uniting global communities, scaling cybersecurity solutions, and measuring their impact. In the five years since its launch, GCA has grown to include over 150 organizations as partners, across over thirty countries, and all sectors of the economy. Partner organizations include industry, governments, academia, and other nonprofit organizations.

Examples of GCA's work include support for Domain-based Message Authentication, Reporting, and Conformance – email security protocols known as DMARC, the development of a protective domain name service (DNS), and the creation of cybersecurity toolkits for at-risk organizations and populations. In supporting DMARC, GCA developed a leader board of domains that have fully implemented the tool, conducted multiple boot camps to train administrators on the proper implementation of DMARC, and translated resources guides into eighteen languages. A 2018 study (Shostack et al., 2018) found that the estimated value to the 1,046 organizations that deployed DMARC at a policy level of "reject" or "quarantine," after using GCA's tool, is likely $19 million (USD).

GCA developed a protective DNS service called Quad9 in collaboration with IBM and Packet Clearing House. Quad9 protects users from accessing known malicious websites by leveraging threat intelligence from multiple industry leaders and blocks an average of over 15 million threats per day for users in over eighty-eight countries. A 2019 study (Shostack et al., 2019) found that the use of DNS firewalls can prevent more than 33 percent of cybersecurity data breaches from occurring.

More recently, GCA combined these projects with free resources from software application developers to develop cybersecurity toolkits for small business, elections administrators, and journalists. The toolkits recommend resources to help these organizations and individuals implement internationally recognized cybersecurity best practices. Each toolkit includes several tools, together with

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Ninety-one percent of cybersecurity incidents begin with a phishing email (FireEye, 2018). In a phishing attack, a malicious actor poses as someone else and sends an email to a victim in order to trick the victim into taking a particular action – often clicking a link that can give the malicious actor account credentials or access to the victim's device. In the absence of multifactor authentication, accounts and devices compromised via phishing or other means can be leveraged for further exploitation. Actors attributed to nation-states have successfully deployed these tactics in a number of high-profile incidents, including the phishing attacks against staff of the Office of Personnel Management (OPM) in 2015, the Democratic National Committee in 2016, and various organizations in 2020.

1 OFFICE OF PERSONNEL MANAGEMENT

In 2015, the global community learned that actors attributed to China were allegedly accessing the email accounts of top US government officials. Also in 2015, information technology staff at the Office of Personnel Management (OPM) discovered that personnel files had been compromised (Fruhlinger, 2020). Among the personnel files that were accessed were approximately 4 million SF-86 forms, which contain extremely personal information, as well as fingerprint records, gathered in background checks for people seeking US government security clearance (Fruhlinger, 2020). After initially obtaining copies of manuals and other network architecture documents the actors moved laterally throughout the network, which had not implemented multifactor authentication. Public reports suggest the actors explored the network for three years before they were discovered and that the incident affected more than 21.5 million individuals (Starks, 2016).

Further exacerbating the initial breach, after the OPM discovered the compromise, it offered employees a credit and identity protection plan. Almost immediately after OPM sent email notifications to register for their credit monitoring services phishing messages appeared (Vaughan-Nichols, 2015). Malicious actors with knowledge of the planned offering leveraged it to obtain account credentials and personal information from OPM staff. While some staff did login and gave the actors access to their personal information, others stopped before entering their data. Cybersecurity awareness training is said to have, in part, limited the impact of the credit monitoring phishing campaign (Rein, 2015).

brief overviews of the need for the tool and step-by-step instructions to guide users through the tools' set up. A community forum and learning management system further support users in their use of the resources. The toolkit for small business is available in four languages, and GCA is assessing methods to measure the toolkits' impact.

GCA works to eradicate cyber risk and improve the connected world. GCA projects focus on the most prevalent cyber risks individuals and businesses face by developing and deploying practical solutions that measurably improve the security of the digital ecosystem; GCA offers these resources at no cost to the global community. GCA is dedicated to increasing cyber awareness and hygiene across all layers of society through awareness-raising campaigns and civil society engagement.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

2 DEMOCRATIC NATIONAL COMMITTEE

On March 19, 2016, John Podesta, the then chair of Hillary Clinton's presidential campaign, received an email purporting to be a Google security alert. Podesta clicked on the link and entered his password into a fake Google log on page through which the actors collected his username and password. As a result, the actors gained access to a decade of his emails (Lipton, 2016). Months later, on October 9, WikiLeaks began publishing thousands of Podesta's compromised emails. Subsequently, several cybersecurity firms attributed the attack to a Russian intelligence unit code-named "Fancy Bear," which has been active since the mid-2000s, and is known among other things for its technique of registering domains that closely resemble domains of legitimate organizations they plan to target. Fancy Bear has also been linked publicly to intrusions into the German Bundestag in 2015, among other intrusions.

3 "MUSTANG PANDA"

January 2020 witnessed a surge in registered domains related to the coronavirus, followed by a spike of cyber incidents. According to Recorded Future's report (Gorey, 2020), malicious actors use COVID-19 as phishing lures for malware, and at least three cases have potential links to nation-state actors. Among them, the "Mustang Panda" campaign has alleged ties to a Chinese government-linked group. The lure used in this campaign was a file discussing COVID-19, purporting to be from the Vietnamese prime minister, Nguyen Xuan Phuc. Once opened, a malicious code could take over the system. Additionally, countries such as the United States, Italy, Ukraine, and Iran have been the focus of related phishing attempts. Malicious actors used trusted organizations as lures for their scam emails, such as pretending to be the World Health Organization and US Centers for Disease Control and Prevention. The malicious emails often use language creating a sense of urgency, or attachments, or links that are said to contain additional information.

At least three cyber hygiene resources can prevent or reduce attacks like the three just mentioned. These resources include deploying Domain-based Message Authentication, Reporting, and Conformance (DMARC), using a protective Domain Name System (DNS), and enabling multifactor authentication. None of these resources alone can prevent a significant cyber incident 100 percent of the time, and they do require investment in human capital. Nonetheless, when implemented across the ecosystem they can have a significant impact. At a minimum, their use can force malicious actors to change targets, tactics, techniques, and procedures. By limiting the impact of phishing and the incidents that may follow, the ecosystem can stabilize, which can support cyber peace.

DMARC is an email authentication, policy, and reporting protocol. DMARC builds on the widely deployed SPF and DKIM protocols, adding linkage to the author ("From.") domain name, published policies from recipient handling of

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

authentication failures, and reporting from receivers to senders, to improve and monitor protection of the domain from fraudulent email. DMARC allows the sender to indicate that their messages are protected and tells the receiver what to do if one of the authentication methods passes or fails – either send the message on or reject the message to junk. DMARC also prevents the dissemination of fraudulent email from an organization's domain. DMARC deployment is a public sector requirement in Australia, Canada, Denmark, the Netherlands, the United Kingdom, and the United States. Moreover, beyond good policy, DMARC prevents significant losses to the global economy. A 2018 study found that the estimated value to the 1,046 surveyed organizations that deployed DMARC at a policy level of "reject" or "quarantine" approached $19 million (USD) (Shostack, Jacobs, & Baker, 2018).

The use of multifactor authentication (MFA) provides an additional effective, low-cost barrier to phishing attacks. A recent survey found that 74 percent of breaches were the result of abuse of privileged credentials (Columbus, 2019). Phishing attacks are one technique used to obtain passwords for use in future exploitation. MFA involves the use of a password plus an additional source of validation, such as a one-time token, to verify a user before granting access to an account. Where enabled, MFA can prevent a malicious actor from using a compromised password to access an account or, in the case of OPM, moving practically uninhibited throughout a vast organizational network.

Additionally, configuring a protective DNS on home and organizational routers can help protect Internet-connected devices against malicious activity. A protective DNS prevents access to known malicious domains by not resolving the DNS query. In doing so, the protective DNS prevents access to a range of threats including malware, ransomware, phishing attacks, viruses, malicious sites, and spyware. Furthermore, using a protective DNS can provide organizations with metrics about the health of their networks and can inform organizational, including national level, incident response functions in the event of a successful attack. One such service, Quad9, protects users from accessing known malicious websites by leveraging threat intelligence from multiple industry sources and blocks an average of over 15 million threats per day for users in over 88 countries. A 2019 study found that the use of DNS firewalls can prevent more than 33 percent of cybersecurity data breaches from occurring (Shostack, Jacobs, & Baker, 2019). The UK Cabinet Office has mandated the use of protective DNS by the public sector. The US Cybersecurity and Infrastructure Security Agency (Nyczepir, 2020) and the National Security Agency are also piloting similar services for their communities of interest (Baksh, 2020).

More recently, actors attributed to nation-states have also capitalized on organizations' failure to patch software and backup data to cause unprecedented losses to the global economy. The Wanna Cry and NotPetya cyberattacks are examples of these incidents. In light of these tactics, two additional best practices can further limit the ability of malicious actors, acting on their own behalf or on behalf of nation-states, from using ICTs to destabilize international order.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

4 WANNACRY

In 2017, actors reportedly affiliated with the government of North Korea used ransomware to cripple computer systems around the world (Latto, 2020). The attack was an example of crypto-ransomware, a type of malicious software used by cybercriminals and other actors to extort money. Ransomware accomplishes this by either encrypting valuable files, rendering them unreadable, or by locking the computer, rendering the computer unusable. Like other types of crypto-ransomware, this attack, dubbed WannaCry, took data hostage, promising to return it upon payment of the ransom.

WannaCry began in May 2017 and spread through computers operating Microsoft Windows (Latto, 2020). Users' files were held hostage, and the actors demanded a Bitcoin ransom for their return. The cybercriminals responsible for the attack took advantage of a previously disclosed vulnerability for which a patch was available. Unfortunately, many individuals and organizations had not regularly updated their operating systems and so were left exposed to the attack. The WannaCry ransomware attack impacted approximately 230,000 computers across 150 countries in just one day – many of them belonging to government agencies and hospitals, including thousands of National Health Service (NHS) hospitals and surgery centers across the United Kingdom (Latto, 2020). The attack affected a third of NHS hospitals, with estimated costs of £92 million after 19,000 appointments were canceled as a result of the attack (Field, 2018). Globally, losses due to WannaCry have topped $8 billion USD (Lemos, 2020).

5 NOTPETYA

The 2017 NotPetya attack offers another example of the importance of maintaining up-to-date software. In NotPetya, actors attributed to Russia launched destructive malware adapted from a series of vulnerabilities common to unpatched Windows operating systems. More specifically, they combined the exploit used in WannaCry together with a password harvesting tool called MimiKatz (Greenberg, 2018). By exploiting vulnerabilities in applications in wide use by the private and public sectors, the NotPetya attack quickly spread from targeted Ukrainian banks, payment systems, and federal agencies to power plants, hospitals, and other systems worldwide. Global companies, including Maersk, Merck, and Mondelez, found their systems impacted, with total losses approaching $10 billion USD (Greenberg, 2018). To date, NotPetya is the costliest attack to ever occur. Yet, had the computers been patched, NotPetya likely would have had far less of an impact because it would have had fewer unpatched systems to leverage into patched systems.

Most recently, in September 2020, a woman in Germany reportedly died after the hospital proximate to her was the victim of a ransomware attack, leading to delay in her care. This incident is the first death publicly attributed to a ransomware attack.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Unfortunately, a 2020 study found that 80 percent of observed ransomware attacks in the first half of 2020 used vulnerabilities reported and registered in 2017 and earlier – and more than 20 percent of the attacks used vulnerabilities that were at least seven years old (CheckPoint, 2020). Thus, without a significant shift by key stakeholders within the ecosystem, particularly governments and entities that develop and maintain connected systems, it will likely not be the last.

These ransomware incidents highlight the importance of enabling automatic software updates where appropriate for the operating environment, and otherwise establishing policies for the prioritization and installation of updates. In addition to ensuring software is up to date, appropriately maintained file backups can also mitigate the risk of ransomware. Ransomware targets that maintain clean and timely backups are often able to avoid significant impact from an attack and continue operations without major delays.

6 CONCLUSION

These cases illustrate that the threat from the malicious use of ICTs is real and that known, effective, accessible, and low-cost resources exist to prevent and limit this threat. Still, reducing cybersecurity risk is a continuous process that requires the use of multiple tools together with human capital. Unfortunately failure to employ cyber hygiene collectively has contributed to significant losses globally, including human life. With the increasing, unavoidable dependence on ICTs for everything from governance and economic development to social engagement, inaction becomes increasingly perilous, especially for governments.

Promisingly, an increasing number of national policies are beginning to require the use of cyber hygiene measures in the public sector. This trend reflects a future reality where use of these capabilities is no longer an option, it is the norm. As a result, a state failing to support their implementation may eventually become the cyber equivalent of a safe harbor. Ultimately, despite what society is often led to believe, what stands in the path of cyber peace is not technology, but political will.

REFERENCES

Baksh, M. (2020). NSA piloting secure domain name system service for defense contractors. *Nextgov*. Retrieved October 28, 2020, from www.nextgov.com/cybersecurity/2020/06/nsa-piloting-secure-domain-name-system-service-defense-contractors/166248/

Buchanan, B. (2020). *The hacker and the state: Cyber attacks and the new normal of global politics*. Harvard University Press.

Center for Internet Security. (Unknown). Ransomware: Facts, threats, and countermeasures. Retrieved October 2, 2020, from www.cisecurity.org/blog/ransomware-facts-threats-and-countermeasures/

CheckPoint. (2020). Cyber attack trends, mid-year report. *CheckPoint*. Retrieved October 29, 2020, from www.checkpoint.com/downloads/resources/cyber-attack-trends-report-mid-year-2020.pdf

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Columbus, L. (2019). 74% of data breaches start with privileged credential abuse. *Forbes*. Retrieved April 21, 2020, from www.forbes.com/sites/louiscolumbus/2019/02/26/74-of-data-breaches-start-with-privileged-credential-abuse/#7bd92a33ce45

Field, M. (2018). WannaCry cyber attack cost the NHS 92m as 19,000 appointments cancelled. *The Telegraph*. Retrieved September 22, 2020, from www.telegraph.co.uk/technology/2018/10/11/wannacry-cyber-attack-cost-nhs-92m-19000-appointments-cancelled/

FireEye. (2018). Email Threat Report. Retrieved October 30, 2020, from www.fireeye.com/offers/rpt-email-threat-report.html

Fruhlinger, J. (2020). The OPM hack explained: Bad security practices meet China's Captain America. *CSO*. Retrieved April 21, 2020, from www.csoonline.com/article/3318238/the-opm-hack-explained-bad-security-practices-meet-chinas-captain-america.html

Gorey, C. (2020). National-state actors may be running phishing scams that exploit the coronavirus. Siliconrepublic. Retrieved May 5, 2020, from www.siliconrepublic.com/enterprise/coronavirus-phishing-scams

Greenberg, A. (2018). The untold story of NotPetya, the most devastating cyberattack in history. *Wired*. Retrieved September 22, 2020, from www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

Greenberg, A. (2019). *Sandworm: A new era of cyberwar and the hunt for the Kremlin's most dangerous hackers*. Doubleday.

Latto, N. (2020). What Is WannaCry? *Avast Academy*. Retrieved April 22, 2020, from www.avast.com/c-wannacry

Lemos, R. (2020). Three years after WannaCry, ransomware accelerating while patching still problematic. *DarkReading*. Retrieved October 29, 2020, from www.darkreading.com/attacks-breaches/three-years-after-wannacry-ransomware-accelerating-while-patching-still-problematic/d/d-id/1337794

Lipton, E., Sanger, D., & Shane, S. (2016). The perfect weapon: How Russian cyberpower invaded the U.S. *The New York Times*. Retrieved April 21, 2020, from www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html

Nyczepir, D. (2020). CISA looks to offer a new DNS resolver to civilian agencies and beyond. *FedScoop*. Retrieved October 28, 2020, from www.fedscoop.com/cisa-dns-resolver-recursive/

Rein, L. (2015). Reacting to Chinese hack, the government may not have followed its own cybersecurity rules. *Washington Post*. Retrieved October 30, 2020, from www.washingtonpost.com/news/federal-eye/wp/2015/06/18/reacting-to-chinese-hack-the-government-may-not-have-followed-its-own-cybersecurity-rules/

Shostack, A., Jacobs, J., & Baker, W. (2018). Measuring the impact of DMARC's part in preventing business email compromise. Global Cyber Alliance. Retrieved September 22, 2020, from www.globalcyberalliance.org/wp-content/uploads/GCA-DMARC-Exec-Summary.pdf

Shostack, A., Jacobs, J., & Baker, W. (2019). The economic value of DNS security. Global Cyber Alliance. Retrieved September 22, 2020, from www.globalcyberalliance.org/wp-content/uploads/GCA-DNS-Exec-Summary-Report.pdf

Starks, T. (2016). House report: Massive OPM breaches a 'failure' of leadership. *Politico*. Retrieved October 29, 2020, from www.politico.com/story/2016/09/opm-cyber-hacks-house-report-227817

Vaughan-Nichols, S. (2015). Phishing e-mail delays OPM hack remediation efforts. *ZDNet*. Retrieved April 22, 2020, from www.zdnet.com/article/phishing-e-mail-temporarily-stops-opm-hack-remediation-efforts/

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

14

Crowdsourcing Cyber Peace and Cybersecurity

Vineet Kumar

1 INTRODUCTION

The Internet's potential can help people from across the globe collaborate and share information for a common cause. Every year, tens of millions more individuals and businesses join cyberspace. However, this newfound access brings in its own set of vulnerabilities, threats, and risks. Crowdsourcing is one way to address these risks by using a systematic approach that makes use of the capabilities of the Internet and its users. When vital information and valuable expertise are shared between people and organizations using crowdsourcing for cybersecurity purposes, it can bring forth positive results for the benefit of all. The CyberPeace Corps is one such crowdsourcing initiative tapping into the skills, expertise, and passions of individuals and groups from all backgrounds to establish cyber peace by collectively building resilience against cybercrime and cyber threats, while upholding the cybersecurity triad of confidentiality, integrity, and availability of digital information resources. Through the crowdsourcing model of the CyberPeace Corps, the idea of a truly global Internet that is trustworthy, secure, inclusive, and sustainable is furthered by leveraging the potential and possibilities of information sharing and collaboration of a large number of people from all over the world.

2 WHAT IS CROWDSOURCING?

The term "crowdsourcing" originates from a collocation of two words – "outsourcing" and "crowd." Simply put, crowdsourcing concerns obtaining information, seeking opinions, and getting the work done with the help of many people who submit data using the Internet as a medium, using the various tools available, such as social media and smartphone apps (Hargrave 2019). People involved in crowdsourcing can be paid freelancers or those who work voluntarily. There are a lot of processes that can take place. However, six identified forms are as follows:

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

1. Crowd innovation
2. Crowd funding
3. Crowd voting
4. Crowd creativity
5. Crowd collective knowledge
6. Micro working (Hargrave, 2019)

A simple example of crowdsourcing would be a traffic app that encourages drivers to report traffic jams or accidents, thereby providing real-time, updated information to other app users. This allows people to save time, take the correct route and, most importantly, be safe during their journey. Some of the crowdsourcing benefits are as follows:

- A wider talent pool is available for getting the work done and can contribute to the cause.
- People can work virtually from anywhere, allowing them the flexibility to choose the location and type of work.
- Various enterprises of different resources and interests can tap into an enormous array of skills, resources, and expertise without incurring significant overheads.
- It also enables businesses to raise a large capital pool for special projects.

The point is that crowdsourcing involves breaking down a complicated project into small achievable tasks that a crowd of people can individually work on to achieve set objectives.

3 CROWDSOURCING CYBERSECURITY

The crowdsourcing cyber conflict model is not a new concept. The 2007 Estonia incident, one of the first and most notable DDoS (Distributed Denial-of-Service) attacks in history, is still fresh in cybersecurity circles (McGuinness, 2017). Malicious actors crowdsourced a series of massive attacks on the Estonian infrastructure, paralyzing the entire city, its largest banking network, and the Parliament (McGuinness, 2017).

But what is interesting is the way Estonia established a model of the first of its kind volunteer cyber force, called the Defence League Cyber Unit (CDL), in 2010. The unit is part of the military in Estonia, but is essentially a civil body with members of the public and private sectors enrolling as experts who render support in the times of a cyber crisis. It started as an initiative during the 2007 attacks, but was capitalized on by Estonia through the institutionalization of the volunteer force into a unit within the military. What this goes on to say is that if crowdsourcing can be used to cause cyberattacks of such a massive magnitude, it can prove useful in fighting cybercrimes as well. Among many others, this incident paved the way for conceiving the creation of a CyberPeace Corps model.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

4 WHAT IS THE CYBERPEACE CORPS?

The CyberPeace Corps is a volunteer-driven initiative by the CyberPeace Foundation for building peace in the cyber world. It is a coalition of citizens, experts, and students who volunteer to come together for sustaining cyber peace. The concept continues to evolve, but it involves a "crowd of" diverse people comprising citizens and organizations who converge as working groups or individual volunteers to foster cyber peace in marginalized communities, organizations, and nations around the world. Currently, over 1,200 CyberPeace Corps members are spread across forty countries are working to enhance their technical capacity against cybercrimes and threats using various modes of communication like social media, street theaters, workshops, webinars, and so on among communities at national and global levels. They also provide support in data collection and analysis to back the training modules developed for workshops, detecting cyberattacks, using machine learning for investigative analysis, and even assist in content creation and dissemination among other activities. The CyberPeace Corps works mainly across four verticals including: Inclusion & Outreach, Collaboration & Connect, Policy & Advocacy, and Innovation & Outreach related to all aspects of cyber peace and cybersecurity. The CyberPeace Corps focuses on the collaboration of people, even from nontechnical backgrounds to build a resilient, safe, and sustainable cyberspace. The CyberPeace Foundation conducts training program for all volunteers who join the CyberPeace Corps and makes them sign a ten-part oath to promote values to ensure peace in cyberspace and strive hard to achieve them. Imagine what malicious actors will do if they manage to access confidential and sensitive information about a country's defense organization. The consequences of such a scenario could be devastating, not only for the organization but also for the common people. On the contrary, suppose a law-abiding citizen gets information about a planned terrorist attack in the country – s/he would report the same to the law enforcement agencies.

Here lies the benefit of crowdsourcing. Crowdsourcing helps create a faceless army of volunteers who can play a stellar role in protecting society from harm. Looking at the scenario from a cybersecurity angle, it is a massive army that helps fight cyber threats on a large scale. The CyberPeace Corps works on this concept of encouraging people to volunteer and fight cybercrime for the good of all. The CyberPeace Foundation has also been working in building a model for children, as well by establishing CyberPeace Clubs in schools. Under the guidance of faculty, school administration, and team at CyberPeace Foundation, students are trained in conducting sessions on cyber safety and also have a continuous dialogue with other students on a resilient and safe cyberspace.

5 HOW CAN CROWDSOURCING WORK IN CYBERSECURITY?

Authorities, police, and agencies have always used crowdsourcing to combat crime in the physical world. The Boston Marathon bombing investigation (Ackerman, 2013)

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

and the Broward County Sheriff case (Contributor BP, 2011) are two prime examples. The public shared media online in large numbers to help the authorities with their investigations. A similar methodology can be used for the purpose of ensuring cybersecurity. There are three prominent fronts where crowdsourcing can help cybersecurity, as discussed below:

- *Collaboration*: People from various locations and different walks of life can put their heads together for a common cause. They can share ideas, work together as a team to bring forth something creative and useful to thwart cybersecurity issues, and offer productive involvement in a cybersecurity project. Synack is an example of one such group of experts ready to respond to organizations' calls and become involved in handling cybersecurity crises so as to combat threats.
- *Sharing Intelligence*: Many experts in the cyber world can contribute vital information to protect numerous people and organizations from serious cyber threats. ThreatExchange, started by Facebook, is an example of an intelligence-sharing platform.
- *Bounty Programs*: These are experts who are not permanent employees of large organizations such as Microsoft, Google, or Apple, but still help them. They can offer their expertise in troubleshooting various organizations' software products to find serious flaws or bugs that can prove fatal sooner or later once discovered by malicious actors. They research independently to identify zero-day vulnerabilities and share valuable information with the organization, thereby helping it develop a patch program and save millions of dollars in the bargain.

6 CROWDSOURCING RESEARCH

Crowdsourcing encourages people from different walks of life to contribute their ideas, based on their usage and expertise, to ensure that diversity of thought processes are accounted for. The CyberPeace Corps has already been employed in various cyber research projects using this method. Involving the public in research has the following advantages:

- Researchers get a chance to understand the public's perspective by involving them.
- Similarly, the public can improve their scientific understanding by participating in the research programs.
- CyberPeace Corps platforms provide the right opportunities for people who want to discover projects or researchers who wish to create new projects.
- It allows various people from diverse fields to assist in cybersecurity tasks for the Corps without enrolling as permanent workers.

As the CyberPeace Corps is a volunteer-centric organization, it relies on community support and participation. There is tremendous potential in the community yet to

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

be tapped. Because there is a severe shortage of cybersecurity professionals capable of handling cyberattacks and related threats, it becomes more relevant to involve the community in projects. Therefore, an initiative like the CyberPeace Corps can help people to contribute their skills and intelligence for the benefit of all. These initiatives have delivered successful results, thanks to collaborative problem solving. Ideas can virtually spring from anywhere, including people not connected to the organization in any way.

7 SERVING SOCIETY

The CyberPeace Corps can help business entities by educating the organization's employees to maintain cybersecurity hygiene. The Corps volunteers are always on the move from one organization to another to drive home the fact that it pays to maintain cybersecurity hygiene. It could be educating people on maintaining strong passwords and discouraging them from sharing them with others to helping people deal with the aftermath of a cyberattack.

The CyberPeace Corps believes in people maintaining self-discipline when using the Internet for personal or business purposes. A simple sharing of an email id and password is enough to clean out a bank account in no time.

8 CREATING AN IMPACT

The crowdsourcing journey comprises four phases through which a potential volunteer has to pass:

Awareness Phase: The awareness phase is the first phase where an individual discovers the initiative through some channels, such as the Internet, professional networks, or a Corps volunteer.

Consideration Phase: The second phase is the consideration phase, where the volunteer learns more about the initiative and decides whether to be involved in it.

Participation Phase: In this phase, volunteers participate in the research and contribute willingly.

Closing Phase: The final phase is the closing phase, where the initiative's compensation occurs.

The CyberPeace Corps assesses the volunteer's journey through this proven model.

9 CALL FOR ACTION

Today, even government authorities and security agencies have put crowdsourcing into practice. By encouraging crowdsourcing in cybersecurity, talented individuals can showcase their capabilities and help organizations thwart cyberattacks.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

- With the industry facing a dearth of cybersecurity professionals, the CyberPeace Corps provides an ideal opportunity for people with the expertise and interest to volunteer to fight cybercrime.
- Interested individuals are welcome to volunteer toward offering their services to the CyberPeace Corps.
- The exciting aspect is that it is not mandatory to have a technical background to volunteer.
- Even people having nontechnical experience can play a critical role in creating cybersecurity awareness among the public.
- Every individual is capable of contributing to safeguarding cyber peace in some way.

As they say, “Security is everyone’s responsibility,” so here is your chance to connect with the CyberPeace Corps help others to be safe online, and contribute to the greater cause of peace in cyberspace.

REFERENCES

Ackerman, S. Data for the Boston Marathon Investigation Will Be Crowdsourced, April 16, 2013, *Wired*, www.wired.com/2013/04/boston-crowdsourced/

Contributor BP, How a Sheriff Uses His 10,000 Facebook Fans to Solve Crimes, October 31, 2011, Consumerist, https://consumerist.com/2011/10/31/how-sheriff-al-lamberti-uses-his-7200-facebook-fans-to-solve-crimes/

Hargrave, M. Crowdsourcing, July 8, 2019, Investopedia, www.investopedia.com/terms/c/crowdsourcing.asp

McGuinness, D. How a Cyber Attack Transformed Estonia, April 27, 2017, *BBC News*, www.bbc.com/news/39655415

Shackelford, S. The World Needs a Cyber Peace Corps, 2017, *Slate*, https://slate.com/technology/2017/10/the-world-needs-a-cyber-peace-corps.html

Weingarten, D. Born in India, CyberPeace Corps Trains Tacklers of Disinformation, Online Challenges, May 7, 2020, Meritalk, www.meritalk.com/articles/born-in-india-cyber-peace-corps-trains-tacklers-of-disinformation-online-challenges/

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

15

Advanced Persistent Threat Groups Increasingly Destabilize Peace and Security in Cyberspace

Anne-Marie Buzatu[*]

1 INTRODUCTION

Their attacks do not typically result in gruesome pictures nor grab the international headlines in the same way as their physical, armed counterparts, but they may be just as deadly or even more dangerous: Advanced Persistent Threat Groups (APTs) are on the rise and changing the very character of modern international conflict today, with yet to be fully appreciated consequences. Operating in obscurity behind screens where they can largely remain anonymous, and called such fanciful names as "Red Apollo" or "Cosy Bear," little is known for certain about who is manning these groups or to whom their allegiances ultimately lie.[1] Rather, analysts try to piece together this information by identifying patterns in cyberattacks, seeing whether the targets of the attack align with the interests of certain states, as well as finding the occasional digital traces that the groups or their members may have left online. What these groups lack in physical bravado they more than make up for in real-world damaging consequences. The COVID-19 pandemic has only served to further accelerate the global dependence upon technologies, providing APTs more opportunities to wreak international havoc and destabilization.

While not officially acknowledged by states, APTs are allegedly run by or sponsored by states to gain unauthorized access to computer systems of governments or companies, where they remain undetected for an extended period of time and gather information, including sensitive information about defense capabilities and critical infrastructure control systems. Recent times have seen the emergence of nonstate sponsored APT groups carrying out large-scale intrusions into

[*] ICT4Peace is an independent foundation that fosters political discussion and common action to support international and human security in cyberspace. To this end, it researches, identifies, and raises awareness about emerging technology challenges, makes policy recommendations, and delivers capacity-building programs.

[1] More information about the forms APT attacks take place can be found here: https://csrc.nist.gov/glossary/term/advanced_persistent_threat.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

government or commercial network systems, sometimes for criminal/financial gain.[2,3]

The "Solarwinds" attack, discovered in December 2020, vividly illustrates both the damage and the uncertainty these kinds of attacks can cause. Analysts said the attack resembled those in the past, thought to have been carried out by Russian-based APTs "Cosy Bear," also known as "APT29,"[4] but Russia has denied any involvement,[5] and the identity of the attack's author is not known for certain, although it seems fairly sure that it is another government.[6] However, apart from the inability to reliably attribute the attack, the extent of the attack itself, as well as the potential security risks it engendered, are also uncertain. What is known is that US agencies, important for the nation's security, were compromised, including the US departments of Homeland Security, State, Commerce and Treasury, the National Institutes of Health, as well as nuclear programs run by the US Department of Energy and the National Nuclear Security Administration.[7] The lack of clarity regarding the information that was stolen, as well as whether critical systems were compromised, has generated a lot of anxiety about the security of US defense systems, with some experts calling for the United States to strike back at Russia.[8] Clearly, APT attacks are turning the traditional international security and peace paradigm on its head, with commensurate risks to our collective safety and security.

2 KINDS OF ATTACKS

APT attacks generally fall into the following categories:

2.1 *Espionage*

APTs infiltrate computer systems and networks and gather information. Targets typically are governments, companies, or other organizations.

For example, an APT group that seemed to be based in China have reportedly targeted South East Asian government machines since at least November 2018, infecting over 200 government machines and even installing backdoors so that they could easily access machines going forward.[9] Other reports claim that

[2] See, Maloney, Sarah, "What is an Advanced Persistent Threat (APT)," last accessed November 29, 2020.

[3] "Why nation-state cyberattacks must be top of mind for CISOs," TechTarget Network, last accessed November 29, 2020.

[4] "Microsoft Discovers a Second Hacking Team Exploiting SolarWinds Orion Software," *CPO Magazine*, last accessed February 16, 2021.

[5] "SolarWinds software used in multiple hacking attacks: What you need to know," ZDNet, last accessed February 16, 2021.

[6] Ibid.

[7] Ibid.

[8] "Cybersecurity experts say US needs to strike back after Solarwinds hack," CBS News 60 Minutes Overtime, last accessed February 16, 2021.

[9] See, for example, "Dissecting a Chinese APT Targeting South Eastern Asian Government Institutions," Bitdefender Draco Team Whitepaper, last accessed November 29, 2020.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

three state-sponsored APTs operating from Russia and North Korea attempted to break into the computer systems of at least seven prominent companies involved in COVID-19 vaccine research and treatment in order to steal sensitive information.[10]

2.2 *Critical Infrastructure Attacks*

The industrial control systems (ICS) that operate and control critical infrastructure systems have been targeted by APTs, which use sophisticated attacks to deactivate, take over control, or destroy them. These include the ICSs of energy grids, water supply systems, electricity production plants, nuclear installations, and banking and telecommunications systems.

One of the earliest of these kinds of attacks that garnered international attention was the 2010 Stuxnet worm that targeted supervisory control and data acquisition (SCADA) systems, which operate the systems that control large-scale machinery and industrial processes, including energy grids and nuclear installations. In this instance, the Stuxnet worm reportedly ruined nearly one-fifth of Iran's nuclear centrifuges by infecting over 200,000 computers and physically damaging approximately 1,000 machines.[11] While no state officially took responsibility for the attacks, analysts largely believe that groups associated with the United States and Israel were behind them.[12]

In the time since the Stuxnet attack, attacks on important and critical infrastructure control systems have continued and increased. For example, in April of 2020, the command and control systems of Israeli water supply systems were reportedly breached by an APT associated with Iran. However, the Israeli government did not disclose any further information regarding the impact of the attack.[13] Additionally, in February 2021, a water treatment plant in the US state of Florida was attacked by a hacker who managed to break into the water treatment control system and increase the levels of lye in the water from 100 parts per million to 11,100 parts per million, which would have made anyone who drank the water very sick. Fortunately, a water plant operator happened to be looking at the ICS screen and witnessed in real time the changes to the levels, correcting them before the changes contaminated the water. However, the computer security systems of the water plant's ICS were not robust enough in themselves to prevent the damage, meaning that if the operator had not happened to be looking at those levels at that particular moment, the water would have been contaminated.[14]

[10] "See Microsoft says three APTs have Targeted Seven COVID-19 Vaccine Makers," available online at: www.zdnet.com/article/microsoft-says-three-apts-have-targeted-seven-covid-19-vaccine-makers/.

[11] "Sheep dip your removable storage devices to reduce the threat of cyber-attacks," Solutions, last retrieved November 29, 2020.

[12] "Stuxnet was work of U.S. and Israeli experts, officials say," *Washington Post*, last accessed 16.02.2021.

[13] Goud, Naveen, "Israel Water Supply Authority hit by Cyber Attack," Cybersecurity Insiders, last accessed November 29, 2020.

[14] "'Dangerous Stuff': Hackers Tried to Poison Water Supply of Florida Town," *New York Times*, last accessed February 16, 2021.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

2.3 *Interference in the Electoral Processes*

APTs are also putting their skills to work at interfering with and undermining national electoral processes.

For example, the US government Cybersecurity and Infrastructure Security Agency (CISA) issued an alert in October 2020, saying that Iranian APTs were creating fictitious websites as well as posting "fake news" to legitimate media platforms in order to undermine public confidence in election systems, as well as to further divide public opinion.[15] US intelligence agencies also reported that Russia interfered in the 2016 US elections, under the direct orders of Russian President Vladimir Putin who, they say, used "troll farms" to create thousands of fake social media accounts to influence popular opinion.[16]

2.4 *Information System Attacks*

Another kind of attack aims to bring down the networks and computer systems so that they are no longer available online.

For example, an APT allegedly associated with North Korea, known as the "Lazarus Group," reportedly took down the Sony Corporation website in retaliation for its release of the film The Interview, a controversial comedy that portrayed US journalists recruited by the US government to assassinate North Korea leader Kim Jong-Un.[17]

2.5 *Ransomware Attacks*

Recently, there has been a sharp increase in ransomware attacks, or attacks in which an organization's data has been stolen or computer systems rendered unavailable, with attackers demanding they be paid a ransom to return data or restore access. In 2020, these kinds of attacks increased by an estimated 319 percent, with perpetrators bringing in at least $350 million (USD).[18]

The above-mentioned "Lazarus Group" APT has also been blamed for one of the most significant ransomware attacks, known as "Wannacry," which was released in May 2017 and infected around 200,000 computers located in 150 different countries.

[15] "Iranian Advance Persistent Threat Actors Threaten Election-Related Systems," US Cybersecurity & Infrastructure Security Agency Alert (AA20–296B), published October 22, 2020, last accessed November 29, 2020.

[16] Ross Brian, Schwartz Rhonda, Meek James Gordon, "Officials: Master Spy Vladimir Putin Now Directly Linked to US Hacking," ABC News, last accessed November 29, 2020.

[17] Heller, Michael, "Lazarus Group hacker charged in WannaCry," Sony attacks, TechTarget Network, last accessed November 29, 2020.

[18] "Ransomware gangs made at least $350 million in 2020," ZD Net, published February 2, 2021, last accessed February 16, 2021.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Of particular note, the National Health Service (NHS) hospitals in England and Scotland were hit, requiring the NHS to cancel many noncritical procedures and treatments.[19]

In September 2020, in what was possibly the first account of a ransomware attack on a hospital resulted in the death of a patient in Dusseldorf, Germany. Having fallen victim to a ransomware attack, the hospital had to reroute the patient's ambulance to another hospital, during which the patient died. Of note, the attacked hospital was not the intended victim of the attack, as the ransom note was addressed to a nearby university. The attackers stopped the attack once authorities informed them that it had shut down a hospital, however, this came too late to save the victim.[20]

In addition to the serious, even life or death consequences for health, both of these attacks also illustrate another difficulty about cyberattacks; that is, their often indiscriminate nature, infecting systems and devices across the Internet where they find vulnerabilities, and not just the intended target(s).

3 "CYBER HYBRID WARFARE": AN EMERGING THREAT TO CYBER PEACE

The activities and cyberattacks carried out by APTs are changing the character of international conflict today. In the words of Australia's Defense Minister, Linda Reynolds:

> [w]hat is clear now, is that the character of warfare is changing, with more options for pursuing strategic ends just below the threshold of traditional armed conflict – what some experts like to call "grey-zone tactics" or "hybrid warfare."[21]

More worryingly, the nature of these "grey-zone tactics" by and large slip through the cracks of our international legal frameworks, most of which were constructed around underlying assumptions that attacks would be physical or kinetic, and that states had effective territorial control to uphold their international obligations and protect those within their jurisdictions. By contrast, the "cyber hybrid warfare" paradigm thrives in an environment where "cyberattacks" often do not fulfill the requirements of international conventions, in which states do not acknowledge their responsibilities for such acts, and in which private actors can act on behalf of – or in the place of – international legal personalities.

In this new paradigm, all stakeholders can be authors of attacks as well as the victims – oftentimes both state and nonstate actors are injured by the same attacks

[19] "Cyber-attack: Europol says it was unprecedented in scale," BBC News, published 13 May 2017, last accessed November 29, 2020.

[20] Wetsman, Nicole, "Woman dies during a ransomware attack on a German hospital," The Verge, published September 17, 2020, last accessed November 29, 2020.

[21] Dowse, Andrew and Bachmann, Sascha-Dominik, "Explainer: what is 'hybrid warfare' and what is meant by the 'grey zone'?" The Conversation, published June 17, 2019, last accessed November 29, 2020.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

online – and effective defense against such attacks may more likely come from the private sector instead of public security forces. If we are to adapt the international legal order to one that supports cyber peace, this calls for new thinking and innovative approaches. While the scope of this essay is not such as to go in-depth into conceiving such a paradigm, as this has already been done throughout this volume, in constructing this new framework, from our perspective, the following elements should be considered or reconceived.

3.1 *The Adaptation of International Legal Obligations and Norms to the Cyber Frontier*

While it is generally accepted that "international law applies online as well as offline," what is not always clear is what this means in practice within the interconnected, transborder environment of cyberspace. Furthermore, as states have implemented their international law obligations to reflect national cultural contexts and values, how do we reconcile these often different and sometimes incompatible state-specific standards within an interconnected, largely borderless cyberspace?

3.2 *The Notion of "Effective Control"*

Furthermore, what actor has the ability to effectively control online activities? Is it the company who owns an undersea cable that forms part of the Internet's backbone? Is it the company that owns the computer servers and/or the state in which those same servers are located? Is it the APT that has the knowledge to infiltrate and even control state and commercial computer systems? When reconsidering the notion of "effective control," we should look carefully at which actors have the know-how/capability to effectively stop or prevent the kinds of behaviors online that undermine cyber stability and cyber peace. This issue is part and parcel of creating an effective regime to regulate state-sponsored cyber aggression.

3.3 *Responsibility and Accountability v. Protection*

Finally, as actors have the ability to carry out activities anonymously, private actors sometimes pack more cyber power than states, and states do not publicly acknowledge their involvement in many attacks, how do we craft a system in which there is effective responsibility and accountability for online attacks? Perhaps this question should be turned around and be considered from a human security point of view as the CyberPeace Institute suggests,[22] asking how can we best protect the human rights and safety of individual users/netizens online?

[22] "CyberPeace: From Human Experience to Human Responsibility," Medium, last accessed February 16, 2021.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Recent initiatives offer some promising avenues to pursue. For example, the Office of the High Commissioner on Human Rights (OHCHR) B-Tech project, which aims to apply the UN Guiding Principles on Business and Human Rights to the ICT sector, is looking at how to create a "smart mix of measures" by exploring regulatory and policy responses to human rights challenges linked to new technologies.[23] UN Secretary-General António Guterres launched a High-Level Panel on Digital Cooperation, bringing together actors from public and private sectors to advance discussions on improving cooperation in cyber governance, which resulted in the "UN Secretary-General's Roadmap on Digital Cooperation."[24] Both France's Paris Call for Trust and Security in Cyberspace[25] and New Zealand's Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online[26] call on both governments and the ICT commercial sector to join forces in combatting malicious attacks online. Microsoft has been very active in the cyber governance space, launching a number of different initiatives, such as the Cyber Tech Accord,[27] to advance multistakeholder discussions at the international level. My organization ICT4Peace, a CSO, has called on governments to publicly commit to refrain from cyberattacking critical infrastructures[28] – which in principle should extend to the APTs they are affiliated with – and called for the creation of a state peer review mechanism on the order of the Human Rights Council's Universal Periodic Review to provide some oversight and accountability for states' actions online.[29]

All of these initiatives recognize the piecemeal, polycentric, multistakeholder-driven nature of cyberspace, and further that it will take joint efforts and concerted collaborative action toward a goal that is in all of our best interests: A safe and peaceful cyberspace in which all stakeholders can thrive and in which state and human security go forward hand-in-hand.

[23] "Business and Human Rights Technology Project ('B-Tech Project'): Applying the UN Guiding Principles on Business and Human Rights to Digital Technologies," last accessed November 30, 2020.

[24] For more information, see www.un.org/en/digital-cooperation-panel/, last accessed November 30, 2020.

[25] For more information, see pariscall.international/en/, last accessed November 30, 2020.

[26] For more information, see www.christchurchcall.com, last accessed November 30, 2020.

[27] For more information, see cybertechaccord.org, last accessed November 30, 2020.

[28] "Call to Governments to refrain from carrying out offensive cyber operations and cyberattacks against critical infrastructure."

[29] Cyber Peer Review Mechanism.

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Index

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D

Downloaded from https://www.cambridge.org/core. IP address: 3.135.192.76, on 06 Sep 2024 at 14:44:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/8C458021C6FEC398064867A9B5EA938D