Advanced Sciences and Technologies for Security Applications

Ritu Gill
Rebecca Goolsby *Editors*

COVID-19 Disinformation: A Multi-National, Whole of Society Perspective

Springer

Advanced Sciences and Technologies for Security Applications

Indexed by SCOPUS

The series Advanced Sciences and Technologies for Security Applications comprises interdisciplinary research covering the theory, foundations and domain-specific topics pertaining to security. Publications within the series are peer-reviewed monographs and edited works in the areas of:

- biological and chemical threat recognition and detection (e.g., biosensors, aerosols, forensics)
- crisis and disaster management
- terrorism
- cyber security and secure information systems (e.g., encryption, optical and photonic systems)
- traditional and non-traditional security
- energy, food and resource security
- economic security and securitization (including associated infrastructures)
- transnational crime
- human security and health security
- social, political and psychological aspects of security
- recognition and identification (e.g., optical imaging, biometrics, authentication and verification)
- smart surveillance systems
- applications of theoretical frameworks and methodologies (e.g., grounded theory, complexity, network sciences, modelling and simulation)

Together, the high-quality contributions to this series provide a cross-disciplinary overview of forefront research endeavours aiming to make the world a safer place.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors.

More information about this series at https://link.springer.com/bookseries/5540

Ritu Gill · Rebecca Goolsby
Editors

COVID-19 Disinformation: A Multi-National, Whole of Society Perspective

Springer

Editors

Ritu Gill
Defence Research and Development Canada
North York, ON, Canada

Rebecca Goolsby
Arlington, VA, USA

ISSN 1613-5113 ISSN 2363-9466 (electronic)
Advanced Sciences and Technologies for Security Applications
ISBN 978-3-030-94824-5 ISBN 978-3-030-94825-2 (eBook)
https://doi.org/10.1007/978-3-030-94825-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

In a world of significant global challenges, from an increasingly assertive Russia and terrorism to disruptive technologies and the rise of China, the Transatlantic Alliance is more important than ever. NATO has been, is, and will continue to be a unique exercise in human history. We are an alliance of nations based on fundamental common values, with our values an essential part, and the agility to adjust and adapt to ever-changing circumstances key ingredients of NATO's enduring success. This strategic narrative, defined by our values and actions, is the bond, the glue, the foundation that keeps us together. Sustaining and protecting that narrative is an essential aspect of our Alliance.

Our narrative is clear and straightforward: there is no foundation for decent lives, economic activities, or fulfillment of individual or collective dreams and ambitions without shared security. There is no way to build a prosperous, democratic society without ensuring the foundation of security in the broader sense of that term, built on our shared values and freedom. Ensuring this security has been our *raison d'être*.

This is our history.

Still, history is not static. The narrative is not linear. As we are witnessing in these challenging times, history shows NATO responding to difficult situations, as well as pleasant and unpleasant geopolitical surprises. It has done so by an innate yet fragile capacity to adjust, be proactive, and show agility. This is also in the genes, in the DNA of our great organization. Yet overall this history, the essential mission remains, that is, to keep 1 billion citizens living in NATO, allied nations, safe.

This is our story.

As long as history and nations exist, there will be some form of traditional military threat to our security. NATO will need to understand and adapt to these conventional threats. But we also see today a broader definition of security. The pandemic has only accelerated and highlighted this trend. We see a considerable spike in the intensity and scale of hybrid activities, including increased use of disinformation and propaganda. If we look at the world now and into the future, we see security has become more complex, more transversal in this quest for permanent adaptation.

Our unity and our ability to adapt is even more critical as we face a more unpredictable, complex, and rapidly changing sphere of geopolitical conflict. We face more sophisticated cyber-attacks and disinformation, a more assertive Russian, brutal terrorism, nuclear proliferation, disruptive technologies ever. The security impact of climate change, so crucial for own public opinions, and where NATO is embarking energetically in that direction. And, of course, the most critical game-changer in global affairs in many, many decades, or probably centuries, is the shift in the worldwide balance of power with the rise of China.

Such hostile activities seek to undermine our democracies, our institutions, our shared values on which our Alliances is founded. But perhaps more fundamentally, it seeks to undermine the trust of our citizens in the very institutions of democracy because this competition, global competition that never ends, historically, is today also a competition of fundamental ideas. And the most crucial part is what we do, not only to describe this intimidation as worrisome and concerning but to take concrete actions to face this phenomenon. We must continue to invest in our strategic communications capabilities and build an evidence-based understanding of the (dis)information environment to contest a highly contested communications space.

In the last past 18 months only, COVID-19 has shown us how disinformation and propaganda can be used to sow distrust, undermine science, and threaten lives and our security. We have seen dangerous disinformation campaigns about the origins of the Coronavirus, cyber-attacks on hospitals, the nature of Alliance activities, and online espionage on medical research centers working on vaccines and treatments.

Our information environment assessment capability must have well-trained staff, supported by the right technology, with the right scientifically informed understanding to navigate this extraordinarily complex field. It will be essential to develop objective-driven, proactive communications strategies, responding rapidly to the increasing volume of information attacks on our institutions, our people, and our values.

NATO has shown that we remain ready and able to defend our nations, that allies and partners are supporting each other, and that investing in our Armed Forces is an investment in the strength and resilience of our societies. This readiness was demonstrated in no uncertain terms by our Alliance's vital role in fighting this deadly pandemic. This capacity stretches across the maritime, land, air, cyber, and space domains, as well as the information or cognitive sphere.

Taken together, this will help us build the resilience of our societies, and building resilience in our communities requires a whole-of-society approach. Everyone has a role to play; the media, the private sector, academia, civil society, of course, the whole-of-government. But it also includes all the relevant groups in our communities to ensure that our institutions and citizens are equipped for today's digital realities. Beyond the nations that make up the Alliance, we must work in concert with our partners to strengthen our collective ability to prevent crisis and address challenges. It is clear that no nation alone, no continent alone, can cope with the magnitude of the challenges we are now facing as democratic societies.

NATO is about security, but it's also about values, shared values, and shared interests. This Alliance is about freedom, and I think our nations, especially those coming from behind the Iron Curtain, appreciate that strong narrative. To ensure our success and resilience in the face of adversity, we must unambiguously and proactively communicate our core narrative, intent and actions to our publics, our potential rivals, and adversaries.

But the pandemic also showed us that resilience is a whole-of-society concern. But to correctly assess this societal resilience, we need better to understand the nature, diffusion and power of disinformation. With this understanding, we may define even more precise indicators, metrics, and analytics to help Allied nations cope. Leveraging modern research in social and information sciences will support this work. This book will play a small but essential part in building such resilience, representing a whole-of-society perspective including academia, industry, and NATO scientists.

But it also crucial to understand that the speed of innovation and technology implementation and how this might change operations across the operational domains, including the information sphere. Understanding how current and future developments of emerging disruptive technologies (EDTs) will impact these domains of operation will be essential to our operational success. We will need to harness the research, technological and analytical talent in our societies. Bring our universities together, bring our scientists together, bring our top experts, and embed our values and ethical and moral code in our technologies. If we succeed, and I'm very confident that we will, we will maintain our edge and keep our societies safe and resilient.

And that is the concept of NATO. It is not only strong together dedicated to looking forward, introducing innovation, understanding the realities and technologies of today, countering disinformation and propaganda, bringing together the best of our scientific and technological base. I think all our citizens are and will continue to be safer because of that.

This book captures and contributes to the collaborative effort of the nations which make up the Alliance and its partners. It promotes a shared understanding of the information sphere, outlines the attempts to undermine the Alliance during the COVID-19 pandemic, and contextualizes the hazards of disinformation and propaganda. It will guide our travels through and competition for ideas within the information sphere as we evolve as an Alliance.

Dordrecht, The Netherlands

Mircea Geoană, Ph.D.
NATO Deputy Secretary General

Introduction: COVID-19 Disinformation—Implications for Defence and Security[1]

On March 11, 2020, the World Health Organization assessed COVID-19 as a pandemic. Currently, the total number of confirmed cases is over 253 million globally. The pandemic is not only a threat to our health and economy, but also has implications for defense and security. Indeed, defense leaders have highlighted a second fight surrounding the spread of COVID-19, namely disinformation and preparing to face adversaries willing to exploit the public health crisis for nefarious purposes. The current pandemic is a breeding ground for the propagation of disinformation, as it represents the first major global health event in which large social media platforms have become the main distributor of information. With audiences primed to experience anxiety, fear, and confusion surrounding pandemic, malign actors have exploited the health crisis and emotions surrounding the pandemic, turning it into a long-running information war.

Information warfare involves several deceptive and manipulative tactics employed by adversaries to achieve desired goals (Gill et al., 2019). This includes the use of disinformation, which refers to a deceptive technique that is based on the dissemination of untrue information with the *intention* to deceive, manipulate, create division and discord, and sow doubt and fear, versus misinformation in which the disseminator is unaware that the information is untrue and has *no intention* to deceive (Pamment et al., 2018). The goal of disinformation is to divide, foster confusion, and demotivate people. Disinformation is not new, it is part of an old playbook of active measures; however what is new is the development of the internet, more specifically social media platforms, creating a new medium to disseminate messages, expediting the ability to propagate disinformation (Gill et al., 2019). Weaponizing the internet and social media for disinformation purposes, adversaries have an extensive network of internet trolls, bots, and sock puppets to generate and spread disinformation, permeating all areas of the internet most likely to exert influence over people and their perceptions (Gill et al., 2019). Dominant disinformation narratives surrounding COVID-19 reveals that both Russia and China have capitalized on the confusion

[1] The views expressed in this publication are entirely those of the authors and do not necessarily reflect the views, policy, or position of the Government or organization of any given nation.

generated by the pandemic to spread disinformation and shape narratives consistent with their goals of undermining key Western and European Union nations and the North Atlantic Treaty Organization (NATO) (Duncan et al., 2020).

Recognizing the activity of malign actors in the information space in the context of the pandemic, the editors of this volume, who are also the Co-Chairs of the NATO Human Factors and Medicine Research Task Group,[2] pivoted their NATO Group to focus specifically on COVID-19 disinformation, resulting in the current volume documenting the multi-national, whole-of-society research effort conducted in COVID-19 disinformation. As COVID-19 represents the first major health pandemic leveraging the online domain, it is paramount to capture and document the multi-national research and findings from various nations to identify effective counter methods and responses to disinformation. This multi-national volume consists of contributions from Defence Science, academia, and industry. The content is aimed at a diverse audience, including NATO members, researchers from defense and security organizations, academics, and militaries including analysts and practitioners, as well as policymakers.

Chapter 1, "A Political Disinfodemic" identifies the global dominant disinformation narratives and the methods used to spread disinformation. Dr. Kathleen Carley (Carnegie Mellon University, United States) highlights how the online domain served to expeditiously inform the public of the spreading virus, representing the same means through which malign actors propagated disinformation. She estimated there have been hundreds of thousands of distinct pandemic disinformation stories, ranging from drinking bleach to vaccine disinformation. Dr. Carley's chapter focuses on the analysis of Twitter posts, collected since January 2020 in multiple languages, employing machine learning tools, computational linguistic tools, and network science tools to analyze and visualize the data. The author notes that in her analysis, a significant portion of disinformation was propagated by authentic human behavior versus automation (e.g., bots), although automation did contribute to the spread of disinformation. Social media platforms were noted to take action against disinformation and new laws regulating disinformation came into effect; however, the author notes that a typical user has eight social media platforms, resulting in relative ease to move disinformation between platforms. Dr. Carley identifies methods to counter disinformation, noting that many of the attempts failed.

Mr. Dale Reding and Dr. Bryan Wells (NATO Chief Scientist, NATO Headquarters, Belgium) examine COVID-19 disinformation within the broader context of conflict in the cognitive domain from a NATO perspective in Chapter 2, "Cognitive Warfare: NATO, COVID-19 and the Impact of Emerging and Disruptive Technologies". The authors detail why countering disinformation, and more broadly, cognitive warfare is critical to the NATO Alliance writ large, highlighting the disinformation campaign experienced by NATO during the COVID crisis. Linkages to

[2] Specifically the Science and Technology Organization Human Factors and Medicine Panel Research Task Group 293 (NATO STO HFM RTG-293) on Digital and Social Media Assessment for Effective Communication and Cyber Diplomacy; this international research group consists of various nations including Canada, United States, Norway, The Netherlands, United Kingdom and Singapore.

broader strategic disinformation campaigns are outlined, and specifics discussed in the context of state and non-state actor cognitive and narrative competition. Actions taken within the wide NATO S&T technology enterprise are outlined, particularly those related to understanding and countering COVID-19 disinformation. The authors also identify lessons learned from NATO's response to the pandemic. In addition, they examine the current and future role of emerging and disruptive technologies and the scientific challenges they present in the context of cognitive conflict.

In Chapter 3, "Developing Approaches to Detect and Mitigate COVID-19 Misinfodemic in Social Networks for Proactive Policymaking", Dr. Nitin Agarwal and colleagues (University of Arkansas—Little Rock, United States) present novel multimethod socio-computational approaches to analyze disinformation and actors on online social networks during the initial months after COVID-19 was made public. The techniques are presented as case studies in narrative analysis of COVID-19 misinformation themes on blogs. The authors present a COVID-19 misinformation tracker tool developed in collaboration with the Arkansas Office of the Attorney General. The results of this chapter are particularly helpful for policymakers, offering valuable data to make informed decisions about information being propagated and developing appropriate and timely countermeasures. For instance, the authors note that disinformation campaigns frequently lead to physical protests. They highlight a case study from the COVID-19 anti-lockdown protests in the United States demonstrating that action from these communities is more cohesive and thus more successful in their protest efforts. Identifying these events prior to reaching critical mass will allow policymakers to engage in appropriate and timely countermeasures.

Chapter 4, "COVID-19 Disinformation, Misinformation and Malinformation During the Pandemic Infodemic: A View from the United Kingdom" focuses on the disinformation, misinformation, and malinformation propagated by a range of threat actors in order to support their operational, geopolitical, and strategic objectives. Dr. Neil Verrall (Defence Science and Technology Laboratory, United Kingdom) highlights the dominant disinformation narratives propagated by actors, including Russia and China, as well as violent extremist organizations. The author discusses vaccine hesitancy and describes Russia's attempt to discredit the United Kingdom vaccine. The author highlights how the United Kingdom government worked closely with social media platforms throughout the pandemic to help them identify and take action to remove incorrect claims about the virus. The author further describes how disinformation, misinformation, and malinformation may be viewed as a capability or system versus an isolated tactic, highlighting how disinformation may evolve given the future developments in technology and society, including Artificial Intelligence, machine learning, and digital and media lawfare.

An important analysis was conducted by the Netherlands, examining the psychological effects of COVID-19 disinformation in Chapter 5, "Web of Lies: Mapping the Narratives, Effects and Amplifiers of Russian COVID-19 Disinformation". Dr. Aiden Hoyle and colleagues (Netherlands Organisation for Applied Scientific Research, The Netherlands) employed Russian COVID-19 disinformation combined with network methodologies to contextualize a novel hypothetical model of this process. Findings revealed how hostile anti-Western narratives primarily targeted the emotions of anger, disgust, and confusion with the aim to undermine citizens' trust in (supra-) governmental institutions and the media. Specifically, the most targeted types of trust were trust in government institutions and trust in COVID-19 information, with the co-occurrence network showing these were linked to narratives negatively portraying Western political actors and undermining COVID-19 data. The authors indicate this research can aid media practitioners develop interventions, as well as support policymakers bolster societal resilience to hostile disinformation campaigns.

In Chapter 6, "The Asian COVID-19 Infodemic on Instant Messaging Platforms", Dr. Khoo and colleagues (DSO National Laboratories and National University of Singapore, Singapore) examine COVID-19 disinformation on instant messaging platforms. The authors provide background on the use of instant messaging platforms in Asia and discuss how these platforms are a dangerous breeding ground for misinformation. Through a content analysis of COVID-19 claims circulated via instant messaging platforms, the authors identified the dominant false narratives, and multimedia (image, video, or audio) misinformation. They review the current legislative and non-legislative initiatives that discouraged the creation and propagation of false COVID-19 information, as well as the growing body of psychological research on message cues, cognitive vulnerabilities, and personality predispositions that make people susceptible to false information, with the aim to contribute toward a comprehensive strategy to combat misinformation over the longer-term.

Importantly, in Chapter 7, "How to Defend Against COVID Related Disinformation", Mr. Jakub Kalensky (Atlantic Council's Digital Forensic Research Laboratory) focuses on various defensive countermeasures that can be applied against disinformation. These countermeasures fall into four overarching categories, including documenting the threat, raising awareness of the threat, repairing the weaknesses within the information system, and punishing the information aggressors. These categories focus on the desired outcome of the countermeasure, rather than the actor(s) intended to implement them, so as to account for the differences among various countries in the Euro-Atlantic space. The author states that disinformation surrounding COVID-19 should be viewed as a campaign that builds upon significant disinformation efforts by the same actors many years prior versus an isolated event. He further indicates that countering disinformation needs a more holistic approach.

In Chapter 8, "Are You Seeing What I am Seeing? Ensuring Data Relevance for Online Information Environment Assessments", Dr. Arild Bergh (Norwegian Defence Research Establishment, Norway) explores the problems governments need to understand and respond to disinformation, propaganda, and other influence efforts. In order to develop a response, governments need to have appropriate, adequate data from social media. Many must purchase that data from third-party providers. He outlines the obstacles of governments in acquiring the necessary data to track disinformation and deliver their own message to their national audiences. Norway's response to the COVID-19 "infodemic" of disinformation, propaganda, and other influence techniques required this kind of acquisition. Dr. Bergh outlines the obstacles that the Norwegian Defense Research Establishment (FFI) faced in evaluating commercial sources for COVID-19 relevant social media during the "infodemic", the cascade of high volume, high velocity social media content rife with disinformation and propaganda, in early 2020. He explores those obstacles in detail, explaining the legal constraints and the challenges in evaluating vendors and their products. This chapter provides a checklist for data acquisition from third-party vendors for governments seeking to improve their response to disinformation in situations of crisis.

North York, Canada
Dr. Ritu Gill, Ph.D.
ritu.gill@forces.gc.ca

Arlington, USA
Dr. Rebecca Goolsby, Ph.D.
rebecca.goolsby@onr.ca

References

Duncan, M., Fitzpatrick, M., Gill, R., & Waldman, S. (2020). *Coronavirus and disinformation: Narratives, counter-strategies and inoculation of audiences.* Defence Research and Development Canada.

Gill, R., van de Kuijt, J., Rosell, M., & Johansson, R. (2019). *Disinformation in the cyber domain: Detection, impact, and counter-strategies.* Conference Proceedings of the 24th International Command and Control Research and Technology Symposium.

Pamment, J., Nothhaft, H., Agardh-Twetman & Fjallhed, A. (2018). *Countering information influence activities: The state of the art*. Department of Strategic Communication, Lund University.

Contents

Editors and Contributors

About the Editors

Ritu Gill has a Ph.D. in Social Psychology from Carleton University and is currently a Section Head with Defence Research and Development Canada (DRDC). She started her career as a Research Manager in the Research Branch of Correctional Service Canada in Ottawa, and joined Defence Research & Development Canada (DRDC) as a Defence Scientist in 2007. Her research examines online influence activities, specifically, how the internet and social media influences the information environment, including the analysis of online audiences, and how deception techniques employed by adversaries, such as disinformation, impacts audiences. She serves on international defence research collaborations as the Canadian Project Officer for the Tri-lateral Partnership Agreement with Sweden and Netherlands on 'Understanding Influence', and is Co-Lead for the NATO Human Factors and Medicine Research Task Group 'Digital & Social Media Assessment for Effective Communication and Cyber Diplomacy'. In May 2020, Dr. Gill was interviewed by NATO TV Channel identifying dominant Coronavirus narratives propagated by adversaries, potential counter-strategies, and identifying the activities of the NATO HFM research Panel activities.

Rebecca Goolsby (Ph.D.) is a program officer overseeing the program, "Social Networks and Computational Social Science" at the Office of Naval Research. She is a well-known digital anthropologist with a strong background in the study of information warfare, cyberdiplomacy and the use of social media in crisis and disaster. She is a recipient of a Fulbright Award and other honors for scholarship and service in the federal government. She is a highly cited authority on conflict in the information environment with several groundbreaking publications in the field. Her article, "On Cybersecurity, Crowdsourcing and Social Cyber-Attack, published in 2012 by the Woodrow Wilson Center for International Scholarship, introduced the problem of propaganda, trolling, and coordinated information terrorism, providing the first published description of social cyber-attack using crowdsourcing methods. She is

Co-Lead of a NATO Research Technology Group on cyberdiplomacy and communications that has turned its focus to disinformation and COVID-19. She regularly provides briefs on information maneuvers to counter disinformation, social hysteria propagation and crowd manipulations to US forces drawing from a newly developed digital and social media playbook for government and military communicators. In February 2020, she was the keynote speaker for the "Hacking Democracy: Influence Operations in the Digital Age" event in Oslo, Norway as a guest of the Norwegian Defence Research Establishment (FFI) and the Norwegian Atlantic Committee. She is a long-time scholar of digital culture, disaster communications, disinformation and crowd manipulation in social media.

Contributors

Nitin Agarwal (Ph.D.) is the Jerry L. Maulden-Entergy Chair and Distinguished Professor of Information Science with University of Arkansas, Little Rock, USA. He is the Founding Director of the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS.ualr.edu) research center. He studies social computing, (deviant) behavior modeling, social-cyber forensics, health informatics, data mining, and privacy. His research is funded by the U.S. Army, Navy, Air Force, DARPA, National Science Foundation, Entergy Corporation, and Australian Department of Defense (Strategic Policy and Intelligence Group).

Arild Bergh (Ph.D.) is a senior researcher in the Total Defence division at the Norwegian Defense Research Establishment (FFI). He has a doctorate in sociology from the University of Exeter on the topic of conflict transformation. Prior to that, he worked for 20 years as a database designer, software developer and IT consultant in the UK. These complementary skills are used in research on socio-technical issues that affect the Norwegian total defense. His current research focus on propaganda, disinformation and influence operations in social media. Central to this research is the development of knowledge about how different actors influence the public through social media platforms, the motivation for this, what they achieve and how democracies can detect and handle such manipulation. A particular focus is the affordances of different algorithms and technologies and how they can be manipulated. He has published a number of reports and articles on social media based disinformation and influence operations from a socio-technical perspective. He has also done extensive work on the technical, legal and ethical considerations of large-scale social media data collection. He is currently involved in international NATO collaborations on the detection of, and training to resist, influence and disinformation in the online information environment.

Beatrice Cadet is Scientist Innovator at the Netherlands Organisation for Applied Scientific Research (TNO). Her interests lie in taking a multidisciplinary approach and bringing together the technical and psychological aspects of security and safety threats. She has worked on a various range of topics mostly related to the cyber domain, from threat intelligence to social engineering.

Kathleen M. Carley (H.D. University of Zurich, Ph.D. Harvard, S.B. MIT) is a Professor of Societal Computing, in the School of Computer Science's Institute for Software Research at Carnegie-Mellon University; the Director of the Center for Computational Analysis of Social and Organizational Systems (CASOS), and the Director of the Center for Informed Democracy and Social Cybersecurity (IDeaS) which is the CMU center for disinformation, hate speech and extremism on line. She is also the CEO of Netanomics. Her research interests include applying computational social science, cognitive science, organization science, dynamic network analysis, social network analysis, machine learning, data analytics and text analytics to complex social and organizational problems such as social cybersecurity, disinformation, disease contagion, disaster response, and terrorism. She has led the development of infrastructure tools for analysing large scale dynamic network analytical tools (ORA), computational linguistic tools (NetMapper), bot hunters, and various agent-based simulation systems (Construct).

Shu Jia Chee holds a BA(Hons) in Criminology from the University of Manchester (2017) and is currently a researcher in the Social and Behavioural Science Programme at DSO National Laboratories. Her main research interests involve the use of horizon scanning methods and security perspectives to identify strategic challenges and evolving threats in the Asia-Pacific region. Her current research leverages on the psychology of deviance and theories on cognition, motivation, and group dynamics to detect populations at risk of disinformation and extremist beliefs.

Kate Donoven is a Senior Assistant Attorney General in the Consumer Protection Division of the Office of Arkansas Attorney General Leslie Rutledge. She graduated from the University of Arkansas School of Law, holds two bachelor's degrees and is currently pursuing a certificate in strategic management at Harvard Extension School.

Aiden Hoyle is a Ph.D. candidate at the University of Amsterdam, Netherlands Organisation for Applied Scientific Research (TNO), and the Netherlands Defence Academy. He looks at malign information activities, proposing models of the psychological effects that antagonistic strategic narratives can elicit in the civilian domain and testing these models with empirical research.

Jakub Kalenský is Senior Fellow in the Atlantic Council's Digital Forensic Research Lab where he focuses on pro-Kremlin disinformation campaigns and countermeasures against disinformation. He started his work in countering disinformation in 2015 when he was chosen as one of the founding members of the newly formed East StratCom Task Force in the European Union—a unit for countering "Russia's disinformation campaigns" with the mandate from 28 Heads of States and Governments. Mr. Kalenský was the team's lead on countering disinformation and he was the main founder of the task force's flagship product Disinformation Review, which later evolved into the EUvsDisinfo campaign. In the Atlantic Council, Mr. Kalenský served as the disinformation lead in the Ukrainian Election Task Force led by David Kramer; as editor of the already defunct DisinfoPortal.org (currently joined with the DFRLab's website); as editor of external contributions and author of commentaries for the DFRLab. His work also includes constant interaction with the media, briefings, and trainings for governments (including closed-door sessions during the Atlantic Council's StratCom conferences in DC or DisinfoWeek in Brussels), and speaking at conferences (NATO Engages in London, DisinfoAlert in Tbilisi, StratCom Summit in Prague, OSCE/ODIHR meeting in Warsaw, Lenart Meri Conference, Warsaw Security Forum, OSCE Security Days in Prague). He received his master's degree after studying at the Philosophical Faculty at the Charles University in Prague, graduating in Philosophy and Russian language and literature. He is currently studying in a PhD program at his alma mater; his dissertation aims to cover countermeasures against disinformation.

Bernice L. Z. Khoo (Ph.D.) is the Programme Director of the Social and Behavioural Science Programme at DSO National Laboratories, and a participant in the NATO Science & Technology Working Group on Digital and Social Media Assessment for Effective Communication and Cyber Diplomacy (HFM-RTG-293). She received her Ph.D. in Social Psychology from National University of Singapore (2013) and has published work on attitudes research in science education, as well as intra- and inter-group attitudes and behaviours under threats such as mortality salience. Her current research examines individual differences and group effects on attitudes and behaviours to identify potential vulnerabilities and protective factors against disinformation and terrorism in the digital environment.

Judith van de Kuijt studied Conflict Studies at the Radboud University and Military strategic Studies at the Netherlands Defence Academy. Since 2015 she works as a research scientist and consultant at the Netherlands Organisation for applied scientific research (TNO). She currently works for the department Military Operations, conducting research that supports military and security operations. She has experience in behavioral influence, Information Operations, and Intelligence. Next to her work for TNO she works as a military reservist for the Dutch army.

Esther Mead (Ph.D.) is a postdoctoral fellow at the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS). Her research interests include the applications of various machine learning approaches to modeling online human behavior and social media analysis such as refugee migration, the spread of misinformation, toxicity assessment and community polarization.

Natalie Pang (Ph.D.) is Senior Lecturer at Communications and New Media Department, Faculty of Arts and Social Sciences as well as Principal Investigator at the Centre for Trusted Internet and Community, both at the National University of Singapore. Her research and teaching are at the intersection of technology and society, and her research projects are organised under the themes of digital citizenship, digital equity & well-being, and digital humanities.

Thomas (Tom) Powell (Ph.D.) is Scientist Specialist at the Netherlands Organisation for Applied Scientific Research (TNO). His research interests focus on using quantitative methods to understand complex phenomena in the fields of intelligence analysis, political communication, information operations and hybrid threats. Dr. Powell was previously an intelligence analyst at the UK Ministry of Defence during which he also deployed in support of the mission in Afghanistan. More recently, he has conducted research in neuroscience, political communication and media effects. His work has been published in several peer-reviewed journals in these fields, such as Cerebral Cortex, Journal of Communication, Media Psychology and Political Communication.

Mr. Dale F. Reding is the Scientific Advisor to the NATO Chief Scientist at NATO HQ in Brussels. In support of the Chief Scientist, he is responsible for the provision of science and technology (S&T) based advice to military and civilian leaders within NATO, including the study of emerging and disruptive technology trends. Before joining NATO in 2019, Mr. Reding enjoyed a 31-year career as a senior public service executive and scientist within the Canadian Department of National Defence. Notably, from 2013–2018, he served as the Director-General responsible for Air Force and Navy S&T research programs. From 2010–2013, he was Director-General of Canada’s centre for military human performance research (DRDC-Toronto). Prior to 2010, Mr. Reding held a variety of scientific management and research appointments. His most notable positions were as the Chief Scientist for Operational Research & Analysis in Ottawa (2008–2010); a Principal Scientist at the NATO C3 Agency in The Hague (NL) (2005–2008); the Director Operational Research & Analysis in Ottawa (2001–2005); and a Senior Scientist at the North American Aerospace Defence Command (NORAD) / United States Space Command (USSPACECOM) in Colorado Springs (USA) (1991–1996). Mr. Reding received his B.Sc. (1981) and M.Sc. (1984) in Theoretical Physics from the University of Saskatchewan. Following this, he undertook additional graduate research in Mathematical Geophysics.

Billy Spann is a Ph.D. Candidate at the University of Arkansas at Little Rock in the Department of Computer and Information Science. He joined the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS) in 2018 where he studies the online behavior of individuals and groups as they use social media, blogs, and other digital platforms to conduct online collective action campaigns. His research interests include social-cyber forensics, identifying powerful information actors and groups, and modeling connective action in online social networks.

Professor Neil Verrall (Ph.D.) is a Senior Principal Psychologist within the Human and Social Science capability of the UK’s Defence Science and Technology Laboratory. He has spent over 20 years applying psychology and behavioural science to many of Defence’s interests and challenges. He is a nationally and internationally recognised expert in information operations and military ‘Influence’ and has spent the last several years reinvigorating Defence’s interest in military deception and contemporary disinformation and misinformation – topics on which he researches, advises, presents and publishes. He currently sits on the UK Government's Counter-Disinformation Working Group, and undertakes operational support, underpinning research and futures thinking on disinformation. As a deployable Defence Scientist he has been the Head of UK Operational Analysis on operations in both Iraq and Afghanistan, as well as conducting numerous deployed research activities. In 2019 he received the UK MOD Chief Scientific Advisor's Commendation for Exemplary Service, for outstanding achievements in behavioural science that saved lives on operations. He holds several degrees in Psychology at bachelors, masters and doctoral levels, and has been awarded a number of visiting research fellowships and academic visitor at the University of Oxford and Royal Holloway University of London. He is a Chartered Psychologist, Fellow of the British Psychological Society, and a qualified Principal Research Officer within the UK's Government Social Research profession. In his spare time he is a Specialist Reserve Officer in the British Army Reserve, providing expert advice and support on deception, disinformation, influence operations, human sciences and social research.

Bryan Wells (Ph.D.) was appointed Chief Scientist by the North Atlantic Treaty Council (NAC) as of 1 July 2019. In this role, he has three major responsibilities. First, he serves as Chair of the NATO Science and Technology Board (STB). In this context, he serves as the STB's representative to the Secretary General and the NAC and is responsible to the STB for the effective coordination of NATO's Science and Technology (S&T) program. Second, he serves as the senior scientific advisor to NATO leadership, ensuring that appropriate and timely S&T based advice is provided to NATO senior decision makers. Finally, he leads the Office of the Chief Scientist at NATO Headquarters. Prior to his appointment as Chief Scientist, Dr. Wells was the UK Ministry of Defence's Head, of S&T Policy, Strategic Research and International Engagement. His responsibilities included the provision of strategic policy advice on the international and research aspects of the Ministry's science and technology programme.

Additionally, he has recently completed a three-year term (2016–2018) as Chair of the European Defence Agency's Research & Technology Steering Board. Dr. Wells joined the UK Ministry of Defence in 1988. He served as Assistant Private Secretary to the Secretary of State for Defence 1989–1992, and has held a range of other posts, including Deputy Director of NATO Policy 1997–1999, and Director of Counter-Proliferation and Arms Control 2002–2008. During 1999–2002 he was on secondment to the Lord Chancellor's Department (now the Department of Justice) as Head of Administrative Justice. Dr. Wells was educated at St Catherine's College, Oxford (1978–85) and Merton College Oxford (1985–1988). He graduated BA(Hons) in Chemistry in 1982 and was awarded a DPhil in 1985. He conducted three years post-Doctorate research at Oxford University as a Junior Research Fellow at Merton College.

Chapter 1
A Political Disinfodemic

Kathleen Mary Carley

Abstract As the pandemic spread globally so too did disinformation within and across social media platforms. The disinformation stories changed to reflect the changing nature of the pandemic, and became increasingly political in nature. Some of the disinformation stories became part of ongoing or new conspiracy theories. Some of the disinformation stories contained lethally wrong information. Many actors used the pandemic as a backdrop against which disinformation was shared to achieve specific political agendas. Fake news sites, bots, and trolls played a strategic role in spreading this disinformation. However, most disinformation was spread by humans. Disinformation was also used to promote hate speech and aggression against various ethnic and minority groups. Attempts to counter disinformation often failed. This process and the coordination behind some campaigns is discussed as are some of the failed countering attempts.

1.1 Pandemic as Backdrop

The COVID-19 pandemic has been unique in many ways. Compared to other viruses that moved beyond a single nation's border—the SARS-CoV-2 virus was more lethal. Compared to previous plagues—information was now carried through the internet often on social media. Powerful digital platforms moved information in seconds throughout the world enabling the World Health Organization and world leaders to quickly inform their publics of what was happening and what precautions to take. These same media, however, also enabled the rapid spread of disinformation.

[1] The term disinformation is used to refer to information that is incorrect and intentionally spread by those knowing it is incorrect. Whereas, the term misinformation is used to refer to information that is incorrect that is spread by those not knowing it is incorrect.

K. M. Carley (✉)
Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: kathleen.carley@cs.cmu.edu

Netanomics, Pittsburgh, PA, USA

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_1

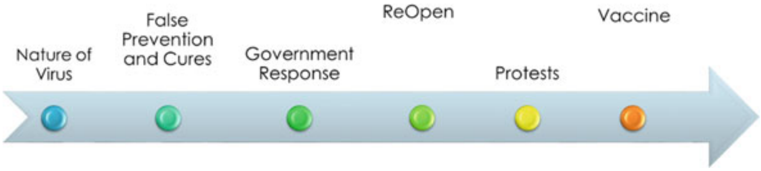


Fig. 1.1 Changing nature of disinformation stories

As the COVID-19 spread throughout the world so too did an epidemic of disinformation and misinformation (Hollowood and Mostrous, 2020).[1] Estimates suggest that there have been hundreds of thousands of distinct disinformation stories with respect to the pandemic. These stories included the innocuous—such as due to the lockdown pollution was lower in Venice and the swans and dolphins returned to the canals. Other stories were lethal—such as drink bleach to cure yourself of COVID-19. Still other disinformation stories were woven together to form larger conspiracy theories—such as Bill Gates invented the SARS-CoV-2 virus and the vaccine, knowing that once the virus was unleashed the world would go on lockdown, requiring all purchases to be made online (cashless society) and he helped create RFID chips that will be injected along with the vaccine so that the government and the Bill Gates of the world can control you via broadcasts to the chip from cellular towers.

The nature of these disinformation stories progressed as the pandemic progressed (see Fig. 1.1). Early on the disinformation was largely about the nature of the virus and its origin. False information about how to self-medicate appeared—often aimed at and spread within specific ethnic-racial communities. As governments started responding, disinformation about what the governments were doing appeared. Further progression led to false information regarding anti-lock down protests. Along with the vaccines came a rash of additional disinformation concerning the vaccines, who got them, and why. Throughout conspiracy theories are being spread, and particularly viral disinformation often gets linked in and becomes part of a conspiracy tale.

As the virus spread, governments responded in diverse ways—and often by framing COVID-19 as a political issue. Pandemic disinformation became increasingly political. Social media participation was manipulated. Disinformation was weaponized. Many organizations began to spread messages on how to separate true from false information regarding the pandemic. For example, UNESCO set up regular educational posts[2] with recommendations on how individuals could inoculate themselves against disinformation such as “avoid sharing anonymous chains.”

Disinformation, however, did not begin with the pandemic. Indeed, disinformation is probably as old as the human race. To be sure there are ancient “written” records of disinformation such as the bas reliefs that Ramses II had carved to celebrate his victory over the Hittites even though that battle was actually a defeat for Egypt.

[2] https://en.unesco.org/news/disinformation-silent-weapon-times-pandemic.

Communication technology has always amped up the ease with which disinformation, misinformation and propaganda can be spread (Posetti & Matthews, 2018). The Gutenberg printing press led to an increase in disinformation—which led to the first large-scale news hoax in 1835—"The Great Moon Hoax" where the New York Sun printed six articles about life on the moon (Thornton, 2000). The pandemic was a perfect breeding ground for disinformation—fear, lack of factual information early on, sufficiently wide spread phenomenon that it could be capitalized on politically, many bored individuals looking for things to do, and so on.

It is important to remember that before COVID-19 became a household word, approaches to stemming the rising tide of disinformation were starting to appear. For example, in 2005 the United Kingdom's Channel 4 News launched a blog with fact checking. By the end of 2010 there were active fact checkers in at least 10 countries (Graves & Cherubini, 2016) with more launching each year. International laws existed and new national laws were emerging to stem the tide of disinformation (e.g., see Baade, 2018; Stray, 2019). There was a growing number of researchers examining disinformation, its cause, spread, adoption, and counters. Thus the pandemic became a turning point for countering disinformation as even social media platforms took unprecedented steps to stop disinformation using tools such as content moderation and promotion of official sites (Butcher, 2021). In this paper, key issues regarding the spread of disinformation regarding COVID-19 are discussed. This discussion draws on a corpus of Twitter posts that we began collecting at the end of January 2020 using keywords in multiple languages for words related to the virus. To analyze this data we set up an analytic pipeline using a number of different tools (Fig. 1.2). These tools included machine learning tools, computational linguistic tools, and network science tools. They were used to do tasks such as identify bots, hate-speech, country and city of origin, and whether the source was a news agency, reporter, or government actor. This was followed by the identification of key actors, key themes, and analysis of who was pushing what stories.

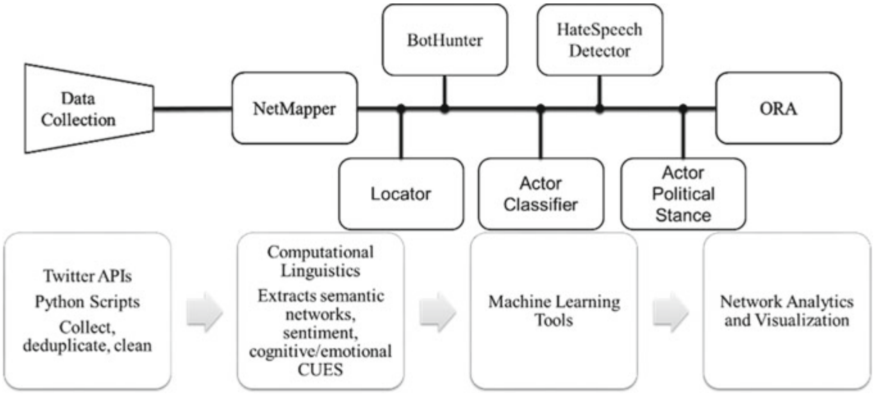

BotHunter
HateSpeech Detector
Data Collection
NetMapper
ORA
Locator
Actor Classifier
Actor Political Stance
Twitter APIs
Python Scripts
Collect, deduplicate, clean
Computational Linguistics
Extracts semantic networks, sentiment, cognitive/emotional CUES
Machine Learning Tools
Network Analytics and Visualization

Fig. 1.2 A data analytic pipeline for assessing Twitter data

Each of the tools in the pipeline has been described elsewhere. NetMapper is a computational linguistic tool that supports extracting semantic networks, sentiment, meta-networks and linguistic cues of affective state from texts (Carley et al., 2018). For the machine learning tools they were retrained and tested on the COVID-19 Twitter corpus to facilitate accurate labeling. These include: BotHunter (Beskow & Carley, 2018), Locator (Huang & Carley, 2017, 2019), actor classifier (Huang & Carley, 2020), hate speech detector (Uyheng & Carley, 2020), actor political stance (Huang & Carley, 2020). ORA is a network analysis and visualization engine with special measures and algorithms for assessing social media data (Altman et al., 2020; Carley, 2017).

1.2 The Good News

As the pandemic spread, news agencies, politicians, academics and many others took up the argument that the world was being over-run by disinformation. But how bad was it? We find that there was a dramatic increase in discussion about COVID-19 and the pandemic. For example, from January 29 to March 4 2020 the number of messages in this space went from just under 3 million a day to over 4 million a day. The rise is even more dramatic as in February the number of tweets per day dropped to under a million as people thought that only China was impacted. Since March there have been approximately 2 million additional users sending messages each day.

Most messages do not contain inaccurate information. Many posts simply were calls for information, comments on strange behavior and shortages, official medical posts provided the most recent understanding. In many other cases, people were recommending home based cures. Though often inaccurate, it is likely these are misinformation. Many messages that contain inaccurate information are either calling it out or appear to be spreading stories that they believe to be true to help. This latter is thought to be happening when people suggest home remedies that are indigenous to a region.

To understand the spread of disinformation, where the sender is trying to deceive, we examined a sample of the tweets with geo-tags. We used these to determine where users were recommending others to go to get additional information; i.e., what URLs were they sharing. Many URLs are to "news" sites; however, the credibility of those sites varies dramatically. Using a five point scale that rates sites from low to high credibility, we identified the fraction of posts that referenced the least credible and the most credible sites. Of the 6,276,440 tweets sampled only 14.38% (902,549) referenced a low credibility or high credibility site. High credibility sites were referenced more than were low credibility sites (see Fig. 1.3). The key is that, proportionally, messages pointing to a low credibility site were more likely to be retweeted than were those pointing to a high credibility sites. In general, we find that posts containing inaccurate information are more retweeted than those containing accurate information. As retweeting is just passing along information, it may be that some of those passing information along did not realize the information was inaccurate.

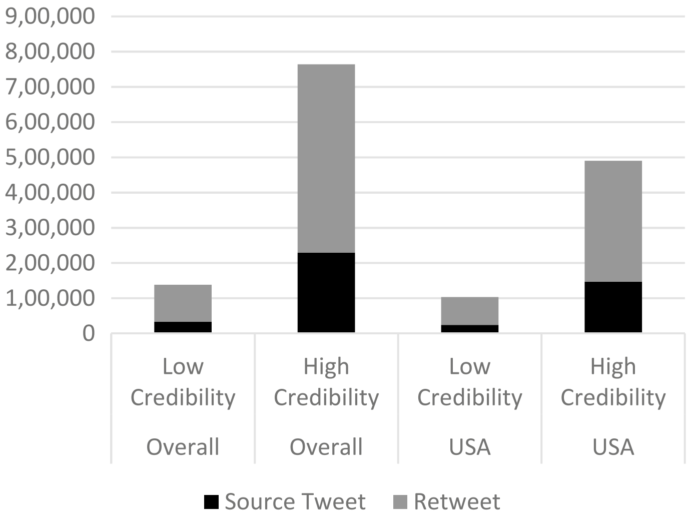


Fig. 1.3 More tweets contained the URLs of credible news sites than low credibility or "fake news" sites

Further, we find that of all the retweets such that the source was in one country and the re-tweeters were in another country, 6.48% of them contain URLs for these "fake news" sites. In other words, retweeting is a critical way in which disinformation moves from one country to the next. However, those tweets containing URLs for credible news sites come from a more diverse set of countries than do tweets containing URLS for low credibility news sites and they end up being retweeted in more countries. This suggests that accurate information is spread more globally; whereas, inaccurate information may be selectively spread to certain countries. It may be that there is a special "trade route" for pushing disinformation to certain countries. In general, the USA appears as a strong consumer and spreader of disinformation. For those tweets that mention a low credibility URL—many appear to come from the USA and to be retweeted mainly by users in Canada and Great Britain (see Fig. 1.4). Whereas, users in the USA appear to be mainly retweeting disinformation first tweeted by those in China, Canada and Great Britain.

Disinformation during COVID-19 is, in a sense, a good news bad news story. Most messages in Twitter do not contain disinformation. A similar point can be made for Facebook, Reddit and YouTube. However, estimates suggest that social media usage increased by 20–87% around the globe during the pandemic (Naeem et al., 2020). The amount of disinformation and the number of new posts per day containing disinformation appears to be at least an order of magnitude higher than in prior crises. Our studies show 4 to a few dozen disinformation stories in many prior elections and natural disasters; but 10's of thousands of stories during the pandemic. For example, Bruno Kessler Foundation (2020) found there were 46,000 new tweets per

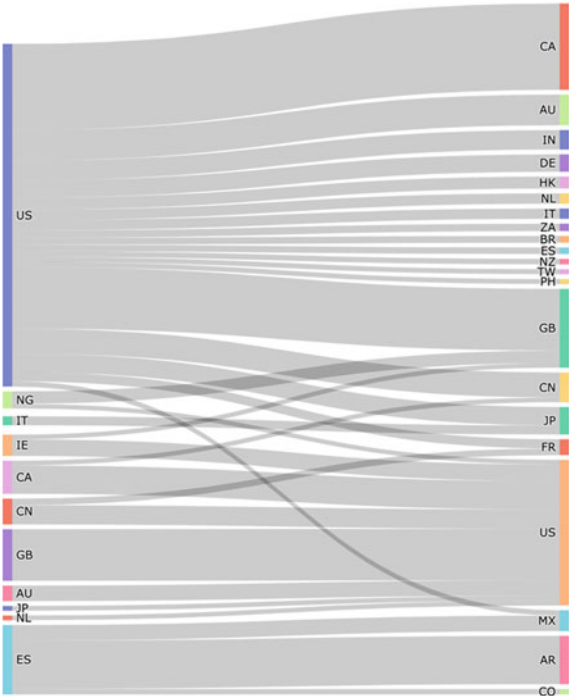


Fig. 1.4 Sankey diagram showing source location and retweet location for tweets containing URLs for low credibility news sites

day containing disinformation. Further, the same disinformation stories are appearing on multiple platforms and spreading from one platform to another.

On the good news side, social media platforms took action and new laws regulating disinformation came into play. However, many of the laws are difficult to enforce. Actions taken by the platforms have not been universally welcomed as: they often appear capricious, may be interfering with free speech, might not be applied in a uniform non-partisan global fashion, may be based on biased algorithms, afford the platforms too much power, and interfere with science. The global nature of social media, countering laws focused on freedom of speech, and technical difficulties in identifying the actual human behind specific posts effectively makes policy and law enforcement difficult. From a platform perspective it is important to note that there is an ecology of platforms. Information or actors banned from one platform may simply migrate to another platform. For disinformation, it often simultaneously exists on multiple platforms—so banning it from one is a temporary measure as it will simply get rewritten, turned into a meme, or combined with other information and then re-appear on the platform that banned it. Since a typical user employs eight social media platforms, it is relatively easy for anyone to move disinformation between platforms.

1.3 Why is Disinformation Spread

Given the life threatening nature of SARS-CoV-2, one might wonder why anyone would spread disinformation—particularly if it is potentially lethal. Analysis of both the stories spread and who was spreading them led us to identify x apparent reasons for spreading disinformation.

As the pandemic began many messages appeared containing inaccurate information about how to prevent or treat COVID-19. Many of these stories involved items that were home remedies for colds or flu—e.g. eat garlic, drink green tea. There appear to be three reasons behind these. First, many people may have been spreading it not knowing it was inaccurate (misinformation) and actually trying to help as these herbal and other home grown remedies may treat symptoms of related diseases. Second, for many ethnic groups, particularly if they were immigrants to a country, the members may have felt that they would not be treated or treated fairly in the official hospital system. Finding remedies by contacting relatives in their old country may have been seen as the only viable medical option. Communicating out this information to others in their community would have been a way of helping. In both these cases we would see this as spreading misinformation; i.e., there was no attempt to deceive. However, the third case, occurred when people knowingly sent message about an inaccurate cure or prevention technique in an effort to harm other groups or mock the overall pandemic. In this latter case, disinformation about cures and preventive techniques was often messaged in a way to attract the attention of particular ethnic, racial or religious minorities by appealing to natural remedies native to their environment. Disinformation, such as the vaccines could harm female fertility, were used to increase vaccine hesitancy in minority groups where there is pressure to conceive children.

Messages such as drink corona beer to get the corona virus, or to prevent it, were sometimes sent for fun. Disinformation in the form of conspiracy theories appear to be sent to recruit new believers, and to reinforce or extend the conspiracy story. Other disinformation appears orchestrated and is part of larger influence campaigns. In some of these cases there are likely to be behind the scene marketers making money by spreading disinformation as has happened in various elections around the world. False images about the size of the re-open protests, disinformation about anti-reopen public proponents, stories about empty hospitals and the like are types of disinformation used to promote social behaviors such as protests, voting, and belief in public institutions. Disinformation about the CDC, WHO, and various medical authorities was used to reduce trust in these public institutions. Disinformation linking different vaccines, social distancing, and facemasks was used to build bridges between diverse groups in attempts to create or strengthen particular movements and political agendas. Such disinformation supported the anti-science movement and a far-right political agenda. Disinformation regarding the actions of governments, police, or leaders such as false descriptions of lock down procedures or their role in promoting a vaccine was used to discredit and so decrease trust in these actors or to bring them undue credit. Some disinformation was aimed at specific minority groups—sometimes to harm and sometimes to keep their voice out of the public debate.

Disinformation was used to fan the flames of hate against various groups; e.g., stories of getting COVID-19 from eating in Chinese restaurants or eating food packaged in China were used to promote hate against those of oriental descent.

In summary, misinformation may be sent in a misplaced effort to help others, strengthen community, or to counter what is felt to be a failure or unfairness of the system. Whereas disinformation was sent in efforts to:

1. Harm others—either a group or an individual
2. Mock an event, person, or issue
3. Have fun or alleviate boredom
4. Recruit people to a cause or belief
5. Bridge or build groups and communities
6. Build an overall story
7. Promote or discourage specific social behaviors such as promoting or voting
8. Discredit or reduce trust in public or private institutions, specific groups or specific individuals
9. Promote or increase trust particular groups or individuals
10. Disrupt civil society
11. Foster a political agenda
12. Reduce the voice of minority groups
13. Promote hate, polarization, and so civil strife
14. Make money.

Importantly, disinformation was used as the pandemic progressed in all these ways. It was targeted at specific individuals, groups, communities and institutions—either to support or discredit them. It was used to impact beliefs and behaviors. It was used to create groups, bridge groups together, recruit members, and to create new opinion leaders. Disinformation was used not just to change what people were talking about, but whom they talked to; not just what they believed, but what communities they joined.

The spread of disinformation and misinformation sometimes led to violence. For example, in India disinformation about Pakistani's tying to spread the SARS-CoV-2 virus led to some attacks on those living in India. Sometimes this disinformation had a financial impact. In the US disinformation that you could get COVID-19 from eating Chinese foods caused many to stop going to Chinese restaurants. Frequently this disinformation had a political impact as it was used to increase partisan divides or credit or discredit various political leaders.

1.4 The Role of Bots

A key issue is assessing who is engaged in the COVID-19 discussion, and so who is spreading the disinformation. In general, for COVID-19 as for most discussions on Twitter, the bulk of the discussion is by “regular users”; i.e., users who are not corporate actors, news agencies, reporters, or government officials. See Fig. 1.5. It is important to note that while news agencies and news reporters are a relatively small fraction of the users, they are among the most influential as it is their tweets that are the most frequently retweeted.

Who are these regular uses? These regular users include people from all walks of life, celebrities, politicians, stay-at-home-parents, teenagers, small companies, doctors, and so forth. In addition, many regular users are bots. A bot is a completely automated account where computer algorithms where computer algorithms are used to determine what to tweet when or which posts to read, like, or retweet. In the overall COVID-19 data that we have collected, between 25 and 40% of the users appear to be bots. These bots contribute a large fraction of the tweets, and an even larger fraction of the retweets.

Of course there are many types of bots. Some bots are simply used as megaphones to retweet messages. Other bots simply tweet out a predefined message; e.g., Big Ben (@big-ben_clock) just tweets the time using “bongs.” Others are used to build groups using sequences of @mentions. Others appear to just be listening. Others engage in chatter. Others spread random messages using key words that the bot director is trying to make trend, and so forth.

It is incorrect to assume that bots are responsible for disinformation. Many bots do not spread disinformation; rather, they may be collating news or broadcasting accurate information.

Disinformation is not created by bots. Rather, disinformation often appears first on blogs, or the news sites with low level credibility. Thus, banning bots will not stop the creation or spread of disinformation.

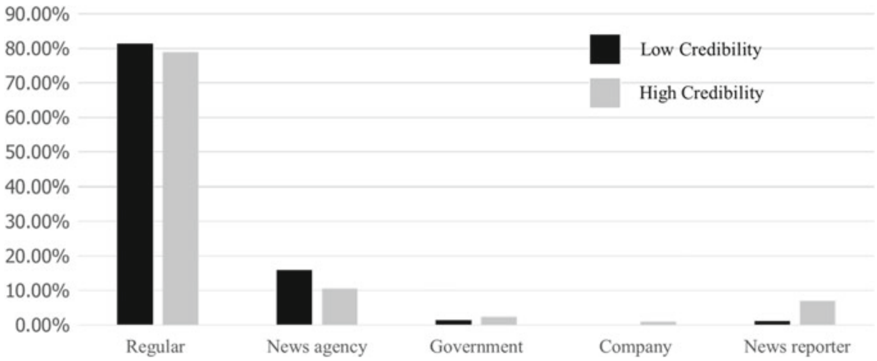


Fig. 1.5 Percentage of users by type

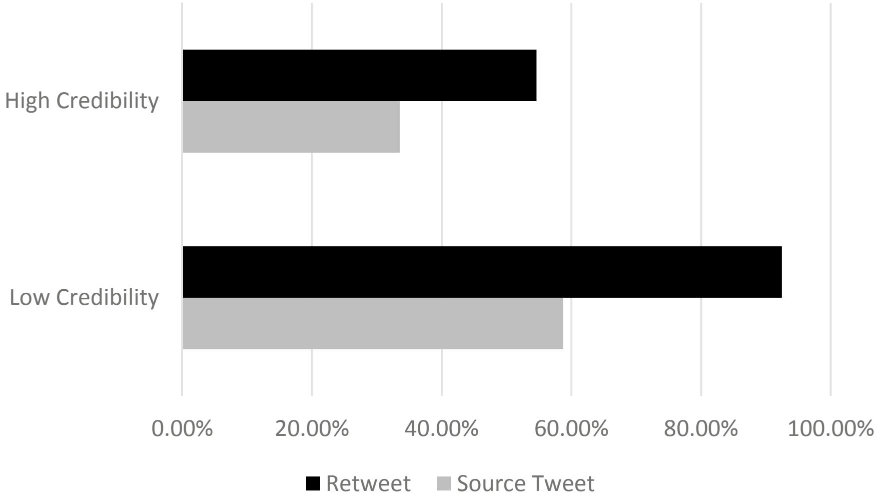


Fig. 1.6 Percentage of tweets and retweets that contain URLs to the lowest and highest credibility news sites that are sent by bots

Bots are often used strategically. Small armies of bots often surround low credibility news sites and will send tweets containing URLs from those sites (see Fig. 1.6). This puts the disinformation from those sites into the twitter verse. Where it is then frequently retweeted by other bots. Such bot megaphones are less common around highly credible news sites. Some conspiracy theories were spread in this way. Indeed as new "news" articles come out, such as those about Gates or 5G, bots are used to link this new information into the ongoing conspiracy.

Small armies of bots surround state sponsored media from China and Russia. In this case the bots are retweeting messages from those sites. That information may or may not be disinformation; but, when it is disinformation these bots are spreading it like a megaphone. Bots are often among the strongest promoters of state-sponsored media; e.g., of the top 50 users who promote Russian state sponsored media, and the top 50 who promote Chinese media—94% are bots. Such bots were used to send stories such as Italians were putting up Russian flags as Russia was more supportive of them as the crisis hit. In point of fact, Russian's did not provide effective aid. In the case of China, these bots spread disinformation such as SARS-CoV-2 was invented in a lab in the US and taken to China. In this way, nation states could engage in information warfare. Common themes were: the origin of the virus, the value or lack of value of a vaccine, the inability of a democracy to protect its citizens, the ability of that nation state to protect its citizen. Many of the stories spread by nation states were in effect propaganda about how well that state was doing, or what good things it was doing for others.

A third strategic use of bots is to discredit critics. This is done by setting up a bot with a name that is extremely similar to the real critic. Then having the bot develop a following through posts like the critic or by having other bots follow it. Then the bot posing as the critic will retweet the state sponsored account disinformation or send posts of support for the nation state. An example of this is CTLau3. The real critic is CTLau2, who routinely sends tweets criticizing China and the Chinese communist party. In contrast the bot, CTLau3, retweeted Chinese state sponsored media, and tweets that contained disinformation about the virus, and tweets that support either China or the Chinese communist party. One such retweet implied that the SARS-CoV-2 virus came from a US bioweapons lab at Fort Detrick.

A fourth strategic use of bots is to build groups either around disinformation, or that can be primed to be receptive to disinformation, or that can be fed disinformation to fan an emotional response such as hate. Following those who follow back, mentioning multiple actors in one message, mentioning multiple actors in conjunction with the same topic—are all ways in which bots can build a group. Once the individuals in the group start following some of the others the group is built and can be manipulated. This approach was used to "introduce" those using hate speech surrounding the spread of COVID-19 to each other. The upshot was communities of hate. In the US some of these were communities of hate directed at those of oriental descent. Disinformation such as eating Chinese food can give you COVID-19 were used to encourage hate and discourage activities such as eating in Chinese restaurants.

Finally bots play key roles in information campaigns aimed at polarizing a community, starting a protest, or orchestrating attacks. In this case collections of bots, new accounts, and trolls are used to create or increase the size of groups, to fuse attention on an issue, to spread messages that emotionally charge the group. During COVID-19 this approach was used to polarize groups against particular ethnic and racial minorities, to encourage the reopen protests, and to strengthen an anti-science movement. A more detailed example is provided later.

1.5 Conspiracies and Lethality

It is tempting to think that disinformation is simply inaccurate facts. If so then automated fact checkers could "solve the problem." However, not all disinformation is spread by inaccurate facts (Fig. 1.7). For example, in spreading anti-vaccine disinformation statements such as "my girlfriend did not get her children vaccinated and they do not have autism." Logical fallacies are often used to spread disinformation. Disinformation is often used by taking images, quotes, and data out of context. Disinformation can be masked by humor, such that the receiver doesn't realize the message was meant as a joke; e.g., drink corona beer to get the corona virus.

Disinformation often takes the form of memes, images, or videos—such as the plandemic video.

Throughout the pandemic, all of these different forms of disinformation were used. Drink bleach or methanol or ethanol to cure COVID-19 were particularly lethal disinformation stories. For example, hundreds in Iran dies from drinking methanol in a desperate bid for a cure (Trew, 2020). On the one hand, these statements to drink

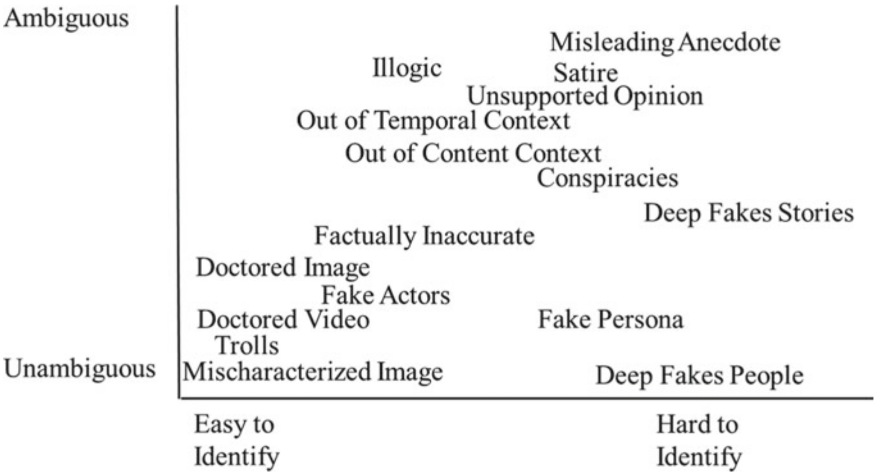


Fig. 1.7 Various types of disinformation that vary in the extent to which they are unambiguously inaccurate and the extent to which they are easy or difficult to identify

these lethal substances are factually inaccurate. However, other ways of spreading disinformation were used to promote them. For example, one meme appeared to show Ina Garten, a celebrity chef, drinking an oversized martini of bleach. In this case satire effectively spread the disinformation, as not all of those seeing the meme recognized it as satire. Groups such as QAnon and their supporters promoted drinking bleach to ward off the virus—talking about it as MMS—a Miracle Mineral Solution (Dickson, 2020). In this case doctored videos were used.

Futurists became concerned about deep fakes. In this case, deep fake refers to an image or video created through the use of deep learning technology applied to massive data sets to create a fake. Deep fakes have, as yet, played little roll in the pandemic. The technology is on the brink of becoming a mainstream way of creating false content. Being machine creatable it is likely to be machine findable.

Deep fake stories played a much greater role in the pandemic. A deep fake story is deep because it is deeply embedded into the culture, history, and existing stories across multiple platforms using a large number of modes such as text, image, sound. For a deep fake story it takes substantial journalistic sleuthing and digital forensics to begin to determine what part of the story is true or false. A classic example are the stories about the white helmets—some of which suggest they are a humanitarian support group, others which suggest they are mercenaries or even terrorists. During the pandemic such stories have taken the form of conspiracy stories. Such stories include the disinformation that COVID-19 is a hoax. In this story, links to stories on low credibility websites, images of "empty" hospitals, arguments that influenza was more lethal, and so forth were used to spread the conspiracy theory. One of the interesting points here is that many of the stories used in arguing that the pandemic was a host were easily identifiable as inaccurate, much of which had been fact checked, and much of which was based on true information taken out of context. For example, the story that influenza was more lethal was an out of temporal context fact as in the number of deaths form COVID-19 early on were lower than for the flu. Stories about empty hospitals included of images of wards in the process of being cleared and prepped for the highly contagious patients with COVID-19.

Throughout history, as crises occur so to do conspiracy theories (Van Prooijen & Douglas, 2017). In 1918 influenza pandemic, for example, one set of conspiracy theorists argued that the flu was a deliberately manufactured bioweapon, while others argued that the Germans were using Bayer aspirin to spread it (Johnson, 2018). What has been different in the COVID-19 pandemic is the ability to spread these conspiracy stories globally, the speed of their spread, and the ability to bring together believers and potentially mobilize them through social media. Conspiracy theories grew in strength and believers in numbers as social media was used to enable collective sensemaking (Kou et al., 2017).

Bots played critical roles in promoting different types of disinformation. For example, bots are more active in promoting lethal and conspiratorial disinformation than in promoting disinformation that is neither lethal nor conspiratorial in nature (Fig. 1.8). We find that as the lethality or conspiratorial nature increases so too does the role of bots—both as the originator and as the retweeter. To illustrate this we consider three different disinformation stories. The first of these is conspiracy in nature and argues that the virus is a human manufactured bioweapon. The second story is lethal and argues that those who get COVID-19 can cure themselves by drinking bleach. The third story is typical of the hoe based cure stories, and argues that eating garlic can served as a preventative and/or cure for the disease. Note in Fig. 1.8, that bots always are more instrumental as retweeters, and that as we move from standard disinformation to lethal to conspiratorial the role of bots increases.

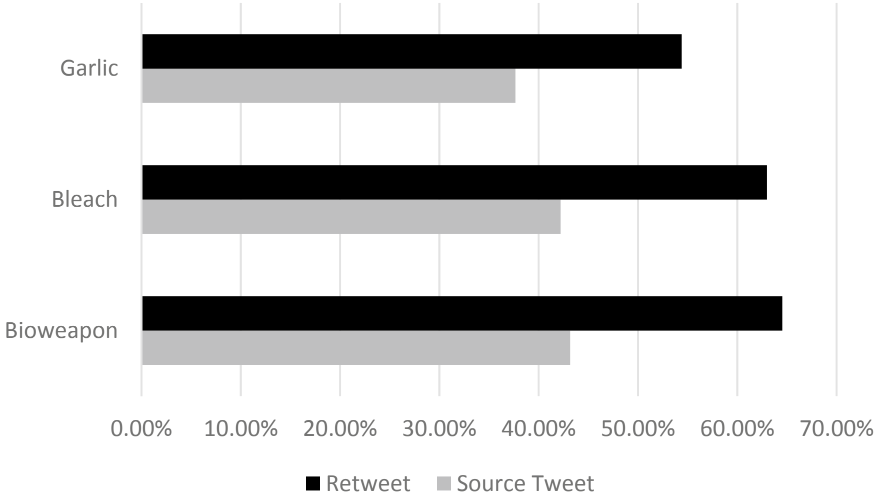


Fig. 1.8 The percentage of source tweets and retweets that reference Garlic as a cure, bleach as a cure, or mention that the SARS-CoV-2 Virus is a manufactured bioweapon that are sent by bots

Table 1.1 Usage of terms related to conspiracies by pro and anti vaccination communities

Terms	Pro-vaccination number of tweets	Anti-vaccination number of tweets
5G, FiveG	2765 (0.17%)	3035 (0.5%)
China, Chinese	38,567 (2.14%)	27,811 (4.7%)
Government & Control	218 (0.01%)	274 (0.04%)

Since many of the tweets about the virus being a bioweapon are found in state sponsored media, the retweets of these are due to the army of bots around such sites.

A key feature of conspiracy theories is that many of them tend to interlink. During the COVID-19 pandemic many different conspiracies and so many different forms of disinformation came to be linked together. To explore this, consider those who are pro and anti- the vaccine. We extracted from the overall twitter corpus a set of messages from users determined to be pro or anti-vaccination in prior studies. These users, based on their tweets, appear to be now also pro or anti the COVID-19 vaccination. We then examined the tweets by these users for messages related to the Bill Gates 5G conspiracy theory, for discussion of government control (as a signal that they were thinking the government was using the pandemic to build controls over the populace), or that were talking about China (generally a signal that they were promoting the ideas that the virus was a bioweapon). We find that the anti-vax community is more likely to mention all three of these conspiracies (Table 1.1).

Although this is a small exploration, it does show a connection among conspiracy theories. It also shows an interesting link to China. In a further examination, we also found that the anti-vax community had more members who followed the state sponsored media accounts from China and/or Russia. In other words, the anti-vax groups are getting fodder for their conspiracies in part from the disinformation coming from foreign states. We also found that there were more bots among the anti-vaccination community than among the pro-vaccination community. And, the bots on the anti-vaccination side often promoted stories about Bill Gates inventing the virus or 5G towers used to control people.

Prior studies have shown that Russian trolls were instrumental in transforming the pro/anti-vax divide into a partisan divide (Walter et al., 2020). They were also instrumental in amplifying the debate (Broniatowski et al., 2018). Our results add to this by showing that the divide is not just due to trolls; but that bots, and regular users are also transforming the nature of the vaccine debate through the use of disinformation. Indeed we find that regular members of the anti-vax group brought in disinformation to the community to foster conspiracies about the effect of vaccinations and the need for them. Conspiracy stories were used to bind individuals into groups and promote vaccine hesitancy. Bots amplified and possibly increased the speed with which this was done by promoting the sane conspiratorial information. Further, we find that links to China, not just to Russia, played a role.

1.6 Orchestrating Change

All of the factors describe herein, came together in a number of efforts aimed at changing society. Two of these will be considered—the reopen campaigns in the US and the anti-science campaigns that have been growing throughout the world. In both cases, the pandemic was used as an opportunity by political actors to try and achieve particular political objectives. The key point, though, is that both of these narratives were part of the same overall orchestrated movement.

The anti-science movement is a complex multi-faceted movement. To be sure, there has always been groups and individuals that were against science. Indeed the anti-science movement can be seen as having its roots in the antireductionist during the scientific revolution (see Toulmin, 1972; Holton, 1993; Jones, 2020 for histories). Russian social media campaigns over the last several years have often been aimed at increasing distrust in science, particularly western science. Russian media fostered the anti-science movement as part of an ongoing campaign associated with anti-vaccination. In the US since 2015 there has been an increasing politicization of science, particularly biomedical science often fomented by Russia (Hotez, 2021).

During the pandemic, though, this somewhat fringe movement gained momentum, became increasingly coordinated, and came into the spotlight. Most of the anti-science focus is aimed at the medical community and argues against preventive measures such as face masks, social distancing, and the vaccine. In early February and March proponents of this movement began to bring together the anti-vaccine groups and coalesce them around being anti a COVID-19 vaccination should one appear. A raft of new accounts were set up many of whom were bots and sent conservative anti-vaccine, anti-science messaging. Increasingly the anti-vaccine groups joined forces with the COVID-19 denial groups. The arguments became increasingly politicized and many came from the extreme far right in countries such as the US and Germany. Various conservative leaders, such as then President Trump, appeared to support the arguments by downplaying the pandemic. Russia encouraged the anti-science rhetoric using bots to promote anti-science messages some of which were coming from Russian state sponsored media. In this way the entire anti-science movement escalated and was now increasingly globally coordinated (Hotez, 2020).

Starting April 5 there was a surge of attacks against expert advisors. In a short period of time there were 45,000 tweets attacking scientists, most in the medical area. In the US many of these were aimed directly at Fauci and Brix. Of those sending the tweets 69% appear to be bots (i.e. have bot-like behavior). In late April there was another surge of new accounts—the majority sent tweets with a conservative stance. Many, though not all, were bots. The new conservative accounts were highly coordinated, sent similar if not identical messages, and were focused on reopening America. Using @mentions ties between these new accounts and existing members of the far right and the anti-vaccination communities bring all of them closer together. Calls for protest went out. Bots were used both to push for protests and make hashtags such as #operationgridlock appear to trend but also to attack democratic governors in the same states. In part his was due to a number of polarization influence campaigns

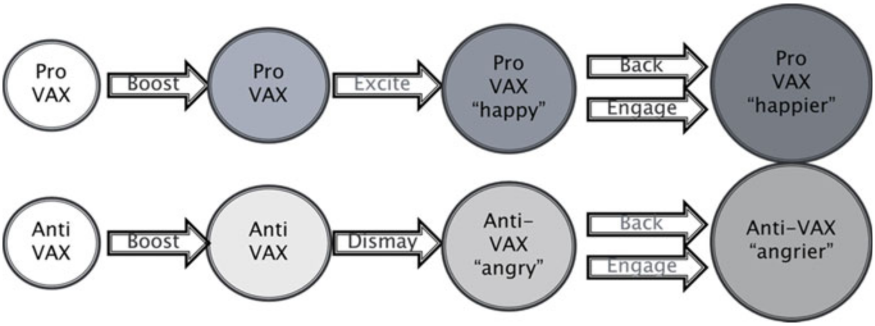


Fig. 1.9 Steps in a polarization campaign

in which the boost, excite, dismay, back and engage maneuvers were used—see Fig. 1.9 (Carley, 2020).

A basic polarization campaign involves identifying an issue where there are strong opinions on both sides—such as pro and anti-vaccination. Then embedding bots and trolls in groups on both sides of the issue, boosting the size of the group and linking the members together so that more will follow or reply to each other. While a bot is a totally automated account, a troll is a human actor generally operating under a pseudonym, disrupting civil discourse, and often attacking a group using expletives. The use of bots and trolls not only increases the size of the group, it also increases the number of messages circulating in the group as the bots and trolls post new messages and retweet others. This increase in activity engages the audience. Such an increase in activity was typically done in response to an external event, such as a leader contracting COVID-19 or the issuance of a new response policy. Messages can be sent to focus the topic around a few issues. One way of doing this is to use similar language another way is to use a small select set of hashtags repeatedly. Messages of excitement on one side should be countered by the bots and trolls with messages of dismay, sadness or fear on the other side and vice versa. The idea is to get the members of each side so agitated that they start outputting their own messages countering the sentiment of the opposing group. Opinion leaders who further serve to narrow the focus and who are more extreme are supported. Messages that add to the focus are used to engage the communities. As the groups grow in both connections among their members and focus on topics the groups become more echo-chamber like in nature and the community as a whole more polarized. Such polarization campaigns were conducted around pro and anti-vaccination, pro and anti-social distancing, pro and anti-facemasks and many other issues.

To promote the protests many such polarization campaigns were linked together; e.g., by forging links among the anti-vaccine, the far-right, the anti-facemask, the pro-reopen, and the covid-is-a-fraud groups. Many of these links were then cemented by appeals to "Making America Great Again" in the US or "National Sozialistisch" in Germany. This was done on only the conservative, anti-vaccine/facemask/social distancing side. Ties among these various "anti-groups" were cemented by appeals to improving the quality of life through herbs and essential oils, and by placing blame on the "non-white" groups. As time progressed more of these groups became linked together. Hate-speech was used to condemn those outside the groups. While the bots did not send original hate speech messages, they did retweet messages containing hate speech within these groups and build bridges among those using hate speech. In some countries this led to new online communities of hate (Uyheng & Carley, 2020). In those countries, within these communities, it was sometimes possible to use bots, trolls and disinformation to change the target of hate. Bonds between the disparate groups were further strengthened by giving them a common enemy—scientists. Science was portrayed as the enemy of the common man, keeping people locked in their homes, needlessly wearing facemasks, and promoting vaccines that could harm. It is important to note that the process of linking these groups together started long before the pandemic (Jaki & DeSmedt, 2019). The pandemic was an opportunity to further this agenda.

In the initial protests, the focus was on the economy and lack of jobs. Some rhetoric argued that the pandemic was not real. Other rhetoric pointed to the idea that you could adequately treat the symptoms of the virus in the same way one might treat the flu with herbal remedies, green tea, and so for the. Attacks on medical experts continued. By the time of the reopen protest in Pennsylvania, there were several new nuances to the discussion. In particular, wearing facemasks and getting vaccinated was transformed from a political stand to a choice and now a right. By framing the argument as a right, with phrases such as "my body my choice" the rhetoric became more powerful and harder to refute. Disinformation, in the form of conspiracy stories, images that made the protests appear bigger than they were, and lies about various political leaders and the medical community, was used. The point here is that there was a concerted campaign to alter the rhetoric so that it became a more aggressive, and so that it became an appeal to values and rights. Messages became simple and clear, and common hashtags were often used.

As discussed then, the orchestration of protests involved a set of complimentary strategies—polarize, building bridges among the conservative side of the polarized groups, and increasing the rhetorical power of the narrative. The story does not stop there. The orchestration also included manipulation of the non-conservative side. As part of the polarization campaigns, bots and trolls were imbedded on both sides of the issue—but they acted in different ways. On the conservative side—they supported the opinion leaders who were extreme, they built bridges among different conservative groups, they sent messages to increase the communicative power of the narrative. On the liberal side—they tried to neutralize the opinion leaders by stopping following, and not retweeting them. They sent messages with no or new hashtags. Many messages were simply random or off topic, and at worst distorted other messages. These dismiss, distort and neutralize tactics obfuscated the key issues, sowed confusion, and effectively discouraged a coordinated response. As confusion was sown among the anti-protest groups, the opinion leaders were attacked, leaving the more liberal side appearing uncoordinated and ineffectual.

1.7 Countering Disinformation

It may seem that disinformation is growing, as are its consequences. Thus a key issue is, can it be countered? To be sure numerous proposals exist such as: promote critical thinking (Lutzke et al., 2019); build automated fact checkers (Thorne & Vlachos, 2018); deception detection (Conroy et al., 2015); new legislation (Marsden et al., 2020; Saurwein & Spencer-Smith, 2020); content moderation (Augustine, 2021); and tagging disinformation (Andersen & Søe, 2020). Indeed many different ways to stop the spread of inaccurate information have been proposed (Cooper, 2020).

Over the course of the pandemic, so far, several different countering strategies have been tried. One strategy was the organic use of humor—such as occurred when people spread images and jokes about corona beer and the virus—making fun of the disinformation that corona beer drinking led to contracting the virus. While not entirely effective, i.e. corona beer sales did go down for a while, overall this tended to stave off further spreading of the disinformation. In other studies we have found that humor including satire can be an effective counter to disinformation; however, it needs to be applied quickly, broadly, and may be more effective on disinformation related to the popular culture (Babcock et al., 2018, 2019). This approach also will not be universally successful as the receiver of a message may not realize it is meant as satire. Thus one person's joke may become another person's truth.

A second strategy was fact checking. Large numbers of fact checking sites throughout the world began to put up information about inaccurate facts. This strategy was used to explain that different cures, different country responses, different claims about hospitals, different claims about whether or not some celebrity had COVID-19 and so forth were correct, partially correct, or incorrect. This helped those who went to fact checking sites. However, most people do not fact check. Fact checking sites in general are often seen as partisan (liberal) (Robertson et al., 2020), and as not being accurate (Brandtzaeg et al., 2018). Fact checking of stories that were less partisan were more likely to stop the spread of that disinformation; however, as the pandemic progressed more and more of the disinformation became partisan. This suggests that the utility of the fact checking began to decrease. However, third parties initiated educational programs to get people to fact check. This approach could be strengthened by the presentation of facts in ways to reduce partisanship, increased training in the use of fact checking, and improved fact checkers.

A third strategy, often followed by the public health and medical community—was to simply try to fill the airwaves with accurate information. This approach was not as effective as it could have been because: (1) at first so little was known about the virus and disease that new messages often contradicted early ones thereby reducing the credibility of the source; (2) accurate information was often presented using more complex language with large number of qualifiers, whereas inaccurate information tended to be presented simply and was more understandable; (3) the sources, such as medical professionals, public health organization, and medical authorities, frequently were understaffed in terms of communications officers and so did not have the resources to promote accurate information in many ways, languages, and across many social media platforms. This strategy could have been strengthened both by higher investment in communication, but also by building support communities in social media around the medical and public health groups so that their messages were passed along. Further, prior studies have shown that when celebrities promote information or disinformation, and when message of excitement or joy are used to promote information or disinformation, that information or disinformation spreads further and persists longer (Babcock et al., 2020). Hence the strategy of filling the airwaves could be strengthened by using celebrities to promote the message and using language in the messages that brings excitement or happiness.

A fourth strategy was tagging—i.e., adding labels to individual accounts or messages that provides the consumer with a warning such as this account is a bot or this message contains inaccurate facts. On the one hand tags make it easier for individuals to choose to engage or not engage with a specific account or messages with a specific tag. The presence of "false" tags, as opposed to "disputed" tags, makes the reader view the message as less accurate (Clayton et al., 2020). However, the impact is stronger if the tag is provided after the readers read the message (debunking), rather than before (pre-bunking) or during the reading of the message (labeling) (Brashier et al., 2021). However, labeling messages can backfire. This was seen to be particularly true when Twitter tagged then President Trump's messages as false (Christenson et al., 2020). This strategy in a sense backfired and appeared to increase polarization and lead to a view by conservatives that the Twitter, the company, was liberally biased. Whether, how, and when tagging may help at this point needs further research.

A fifth strategy was removal—specifically the removal of lethal disinformation and to an extent the banning of those who spread it from some platforms. For example, Twitter removed posts about drinking bleach and put in place a multi-strike rule for banning accounts. This strategy has several consequences. It did keep more people from sharing it online. It also rendered many of the spreaders of that particular disinformation to be unable to spread other disinformation on Twitter as they became banned. However, many of those individuals were members of QAnon and they simply went to other platforms. The disinformation story, though banned by Twitter, still existed on other platforms. YouTube removed the Plandemic video—however it moved to Amazon where it can currently be bought. Removing content and banning actors just made the information and those promoting it, harder to find: it drove them back underground. This strategy made it difficult for scientists to study disinformation and to determine alternative, possibly more effective, counter-measures. Finally, this strategy raises the flag of free speech and civil rights. The value of this is that now in more countries there is ongoing debate about who, if anyone, can censor information and the potential impact of that decision.

Frequently COVID-19 disinformation was countered using multiple strategies. A case in point is the disinformation story telling people to drink bleach (or methanol or ethanol) to cure themselves of COVID-19. While Twitter and a few other platforms censored such stories, other groups tried to counter the same disinformation in a variety of ways. These included pro-messaging campaigns that tried to "bust" the myth by explaining the harms of drinking these substances. Many of these were carried out by the medical community. Another approach was to make fun of the disinformation—showing celebrities drinking the lethal concoction. Fact checking sites put up fact checks calling out the inaccuracy of the claims. Whether these alternative strategies were as or more effective than the platforms removing content and perpetrators is unclear. However, it is likely that the different strategies reached different audiences. In part this is because different groups tend to use different social media platforms. Thus if Twitter, Facebook or YouTube remove content it will only directly impact the individuals on those platforms. However, since people on other platforms talk about what happens on these large platforms, there will be an indirect effect. Putting out positive messaging on Twitter, Facebook and YouTube may directly reach their audiences. Whether such positive messaging spreads to other platforms is unclear.

Collectively, this multi-strategy approach did appear to reduce belief in this disinformation. Was this multi-strategy approach needed? When is a multi-strategy approach needed? To be sure, these questions need more study. However, a few insights can be drawn from the way this disinformation story played out compared to others during the pandemic. A key difference between this "drink bleach" story and the "corona beer" story, both of which were countered with humor—is that statements by the US president seemed to support drinking bleach (celebrity support), bleach appeared to be useful in curbing the spread of the virus (confusing related story). Another key difference between this "drink bleach" story and many of the other stories mentioned is that it was being actively promoted (battleground) by extremist groups often against at-risk minority or ethnic groups (minority groups). As such there was a story battle ground where ideas were competing for the hearts and minds of the populace. More research is needed on what strategies are most effective and how to combine those strategies when these four factors are present—celebrity support, confusing related story, an active idea battleground, and the presence of minority group.

1.8 Conclusion

COVID-19 has been such a dramatic global event that it is easy to think that this event caused disinformation and the anti-science movement. It is tempting to think that the virus caused people to be concerned about how debate on the internet could be safeguarded for all people. However, as we have seen disinformation is old, possibly as old as humankind. The anti-science movement is also old. Social media simply provided a global platform on which humans and machines can be used together to rapidly spread disinformation, to mobilize groups, and to recruit new believers to old conspiracies. COVID-19 rather than being a cause was an opportunity for adversarial actors to pursue their agendas.

A similar case can be made for those studying cyber-communication and developing tools to keep the airways free from undue influence. Prior to the spread of the SARS-Cov-2 virus there were thousands of researchers working in the emerging computational social science area referred to as social cybersecurity. Social cybersecurity is an emerging science focused on enabling online engagements in ways that allow people to communicate and receive news without undue influence or fear of attack (Carley, 2020). The National Academies of Science NAS (2019) has described social cybersecurity as an applied computational social science with two objectives"

- characterize, understand, and forecast cyber-mediated changes in human behavior and in social, cultural, and political outcomes; and
- build a social cyber infrastructure that will allow the essential character of a society to persist in a cyber-mediated information environment that is characterized by changing conditions, actual or imminent social cyberthreats, and cyber-mediated threats."

The thousands of researchers in this space were poised to understand and in some cases support counter-measures to the disinfodemic that spread along with the virus. This paper, is part of this emerging area and build out of that ongoing body of research. For this paper a social cybersecurity technology pipeline, developed pre-pandemic, was used to address the issue of disinformation. Using social cyber-forensic techniques the types of actors spreading disinformation were identified, as were their targets and information maneuvers. Diffusion of disinformation between countries was traced, and dispersion of state sponsored disinformation was discovered.

A pandemic is the kind of crisis event where open honest discourse is needed to stem the tide of fear, increase safety and decrease risk. To be sure, researchers in social cybersecurity have focused on the pandemic, and reached out to help enable improved online engagements during this critical juncture. As a result, new tools such as the hate speech detector and improved automated fact checkers are now appearing. Existing tools have been scale up and improved such as those in the technology pipeline used for this paper. Future work though, needs to better address identifying the motive of those spreading disinformation and strategies for countering a disinfodemic. As was discussed there appear to be many motives for spreading disinformation and many possible avenues for countering. However, the relative prevalence of the different motives and whether the motive affects the impact of messages sent needs further study. As to counter-measures the ones identified are just a few of those that are possible. Research is needed to identify a broader spectrum of counter-measures and to assess their cost and effectiveness relative to various types of disinformation.

Acknowledgements This work was supported in part by the Knight Foundation and the Office of Naval Research (ONR) grants N000141812106 and N000141812108. Additional support was provided by the Center for Computational Analysis of Social and Organizational Systems (CASOS) and the Center for Informed Democracy and Social Cybersecurity. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, ONR, or the U.S. government. The author wishes to thank and acknowledge the contributions to this work of the following individuals while they were Ph.D. students at CMU: Binxuan Huang, David Beskow, Iain Cruickshank, Sumeet Khumar, Aman Tyagi, Joshua Uyheng. The author also wishes to thank and acknowledge the contributions of Matthew Babcock, who was a post-doctoral fellow working with the CASOS group.

References

Altman, N., Carley, K. M., & Reminga, J. (2020). *ORA user's guide 2020* (Technical Report CMU-ISR-20–110). Carnegie Mellon University, School of Computer Science, Institute for Software Research.

Andersen, J., & Søe, S. O. (2020). Communicative actions we live by: The problem with fact-checking, tagging or flagging fake news–the case of Facebook. *European Journal of Communication, 35*(2), 126–139.

Augustine, R. (2021). A Critique on content moderation on Facebook—A study based on 'stop the steel' conspiracy campaign. *Media, Culture and Society, 22*, 24.

Baade, B. (2018). Fake news and international law. *European Journal of International Law, 29*(4), 1357–1376.

Babcock, M., Beskow, D. M., & Carley, K. M. (2018, July). Beaten up on twitter? Exploring fake news and satirical responses during the black panther movie event. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (pp. 97–103). Springer.

Babcock, M., Beskow, D. M., & Carley, K. M. (2019). Different faces of false: The spread and curtailsment of false information in the black panther twitter discussion. *Journal of Data and Information Quality (JDIQ), 11*(4), 1–15.

Babcock, M., Villa-Cox, R., & Carley, K. M. (2020). Pretending positive, pushing false: Comparing captain marvel misinformation campaigns. In *Disinformation, misinformation, and fake news in social media* (pp. 83–94). Springer.

Beskow, D. M., & Carley, K. M. (2018, July). Bot-hunter: A tiered approach to detecting & characterizing automated activity on twitter. In *Conference paper. SBP-BRiMS: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (Vol. 3, p. 3).

Brandtzaeg, P. B., Følstad, A., & Chaparro Domínguez, M. Á. (2018). How journalists and social media users perceive online fact-checking and verification services. *Journalism Practice, 12*(9), 1109–1129.

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5).

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health, 108*(10), 1378–1384.

Butcher, P. (2021). COVID-19 as a turning point in the fight against disinformation. *Nature Electronics, 4*(1), 7–9.

Carley, K. M. (2017). ORA: A toolkit for dynamic network analysis and visualization. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining*. Springer.

Carley, K. M. (2020). BEND: A framework for social cybersecurity. *Future Forces, 6*(2), 20–25.

Carley, L. R., Reminga, J., & Carley, K. M. (2018). Ora & netmapper. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*. Springer.

Christenson, D., Kriner, K., & Kreps, S. (2020, November 25). When Twitter fact-checks Trump's tweets, it polarizs Americans even more, our research finds. *Monkey Cage*. https://www.washingtonpost.com/politics/2020/11/25/when-twitter-fact-checks-trumps-tweets-it-polarizes-americans-even-more-our-research-finds/. Last accessed 3/21/21.

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior, 42*(4), 1073–1095.

Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology, 52*(1), 1–4.

Cooper, M. (2020). 11 Ways to stop misinformation. *IT Now, 62*(2), 26–27.

Dickson, E. (2020, January, 29). QAnon youtubers are telling people to drink bleach to ward off Coronavirus. *Rolling Stone*. https://www.rollingstone.com/culture/culture-news/qanon-conspiracy-theorists-coronavirus-mms-bleach-youtube-twitter-944878/. Last accessed 3/1/21.

Graves, L., & Cherubini, F. (2016). *The rise of fact-checking sites in Europe*. Reuters Institute.

Hollowood, E., & Mostrous, A. (2020, March, 23). Fake news in the time of C-19. *Tortoise*. https://members.tortoisemedia.com/2020/03/23/the-E2-80-90infodemic-E2-80-90fake-E2-80-90news-E2-80-90coronavirus/content.html. Last accessed 2/11/21.

Holton, G. (1993). *Science and anti-science*. Harvard University Press.

Hotez, P. J. (2020). *Anti-science extremism in America: Escalating and globalizing*.

Hotez, P. J. (2021). Anti-science kills: From Soviet embrace of pseudoscience to accelerated attacks on US biomedicine. *PLoS Biology*, *19*(1), e3001068.

Huang, B., & Carley, K. M. (2017). On predicting geolocation of tweets using convolutional neural networks. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (pp. 281–291). Springer.

Huang, B., & Carley, K. M. (2019). A hierarchical location prediction neural network for twitter user geolocation. *arXiv preprint*. arXiv:1910.12941

Huang, B., & Carley, K. M. (2020). Discover your social identity from what you tweet: A content based approach. In *Disinformation, misinformation, and fake news in social media* (pp. 23–37). Springer.

Jaki, S., & De Smedt, T. (2019). Right-wing German hate speech on Twitter: Analysis and automatic detection. *arXiv preprint*. arXiv:1910.07518

Johnson, N. A. (2018). The 1918 flu pandemic and its aftermath. *Evolution, Education and Outreach, 11*(5).

Jones, R. H. (2020). *Reductionism: Analysis and the fullness of reality*. Bucknell University Press.

Kou, Y., Gui, X., Chen, Y., & Pine, K. (2017). Conspiracy talk on social media: Collective sense-making during a public health crisis. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–21.

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, *58*, 101964.

Marsden, C., Meyer, T., & Brown, I. (2020). Platform values and democratic elections: How can the law regulate digital disinformation? *Computer Law & Security Review*, *36*, 105373.

Naeem, S. B., Bhatti, R., & Khan, A. (2020). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*.

National Academies of Sciences, Engineering, and Medicine. (2019). *A decadal survey of the social and behavioral sciences: A research agenda for advancing intelligence analysis*. The National Academies Press.

Posetti, J., & Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation. *International Center for Journalists, 7*, 2018–2107.

Robertson, C. T., Mourão, R. R., & Thorson, E. (2020). Who uses fact-checking sites? The impact of demographics, political antecedents, and media use on fact-checking site awareness, attitudes, and behavior. *The International Journal of Press/Politics, 25*(2), 217–237.

Saurwein, F., & Spencer-Smith, C. (2020). Combating disinformation on social media: Multilevel governance and distributed accountability in Europe. *Digital Journalism, 8*(6), 820–841.

Stray, J. (2019, May). Institutional counter-disinformation strategies in a networked democracy. In *Companion Proceedings of the 2019 World Wide Web Conference* (pp. 1020–1025).

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. *arXiv preprint*. arXiv:1806.07687

Thornton, B. (2000). The moon hoax: Debates about ethics in 1835 New York newspapers. *Journal of Mass Media Ethics 15*(2)*,* pp. 89–100. Accessed 3/18/21.http://www.tandfonline.com/doi/abs/10.1207/S15327728JMME1502_3

Toulmin, S. (1972). The historical background to the anti-science movement: Civilization and science: In *Conflict or Collaboration*, 23–32.

Trew, B. (2020). Coronavirus: Hundreds dead in Iran from drinking methanol amid fake reports it cures disease. *Independent*. https://www.independent.co.uk/news/world/middle-east/iran-coronavirus-methanol-drink-cure-deaths-fake-a9429956.html. Last accessed 2/11/21.

Uyheng, J., & Carley, K. M. (2020). Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines. *Journal of Computational Social Science, 3*(2), 445–468.

Van Prooijen, J. W., & Douglas, K. M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies, 10*(3), 323–333.

Walter, D., Ophir, Y., & Jamieson, K. H. (2020). Russian Twitter accounts and the partisan polarization of vaccine discourse, 2015–2017. *American Journal of Public Health, 110*(5), 718–724.

Chapter 2
Cognitive Warfare: NATO, COVID-19 and the Impact of Emerging and Disruptive Technologies

Dale F. Reding and Bryan Wells

Abstract COVID-19 disinformation campaigns are considered as operations within the cognitive domain. Given its role in hybrid or grey zone warfare, countering disinformation is critical to the NATO Alliance. The disinformation campaign experienced by NATO during the COVID crisis provides an example of low-intensity but high-impact cognitive conflict. This overview of NATO's response to COVID-19 is based on official publicly available NATO materials and the authors' experiences supporting NATO's crisis management and strategic communications during the pandemic. Linkages are made to broader strategic disinformation campaigns and specifics discussed in the context of state and non-state actor cognitive and narrative competition. Actions taken within the general NATO S&T technology enterprise are outlined, especially those related to understanding and countering COVID-19 disinformation. Finally, the potentially destabilising effects of current and future emerging and disruptive technologies are considered as well as the scientific challenges they present in the context of cognitive conflict.

2.1 Introduction

Cognitive warfare, and its older incarnation of information warfare, have evolved considerably over the last hundred years, being named by some as the sixth operational domain for NATO (Bienvenue et al., 2018; Cole & Le Guyader, 2020; Elkins, 2019). Equally, the combination of new technologies and a better understanding of human behaviour has had and will continue to have a substantial role in enabling the increased effectiveness of operations in the cognitive domain and our appreciation of narrative conflict. In addition, the associated disinformation campaigns that have arisen in the context of the COVID-19 pandemic have given new impetus to our need

D. F. Reding (✉) · B. Wells
Office of the Chief Scientist, NATO Headquarters, Boulevard Leopold III, Brussels, Belgium
e-mail: reding.dale@hq.nato.int

B. Wells
e-mail: wells.bryan@hq.nato.int

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_2

to understand, counter and inoculate societies from associated infodemics (Galvão, 2020).

This chapter seeks to place the COVID-19 disinformation campaigns within the broader context of conflict in the cognitive domain from a NATO perspective. First, we explain why countering disinformation, and more broadly, cognitive warfare, is critical to the NATO Alliance writ large (especially in the context of grey zone warfare). Second, to place this in more concrete terms, we outline the disinformation campaign experienced by NATO during the COVID crisis. This overview of NATO's response to COVID-19 is based on official publicly available NATO materials (NATO, 2020b; Ozawa, 2020) and the authors' experiences supporting NATO's crisis management and strategic communications during the pandemic. Third, linkages to broader strategic disinformation campaigns are outlined, and specifics discussed in the context of state and non-state actor cognitive and narrative competition. Fourth, actions taken within the wide NATO S&T technology enterprise are outlined, especially those related to understanding and countering COVID-19 disinformation. Finally, we consider the current and future role of emerging and disruptive technologies (Reding & Eaton, 2020) and the scientific challenges they present in the context of cognitive conflict.

2.2 NATO and Disinformation

As NATO moves towards the future, it finds itself confronting the challenges of a changing world. As noted in the NATO2030 process (Maizière et al., 2020; NATO, 2021a), NATO will need to address a host of returning and emergent challenges such as Russia's aggressive actions, the threat of terrorism, cyber-attacks, emerging and disruptive technologies, the security impact of climate change, and the rise of China. The solutions to these challenges require a delicate blend of political and military action married with the deep collective wisdom contained within the Alliance. Within the three core missions specified by NATO's 2010 strategic concept (collective defence, crisis management and cooperative security) (NATO, 2020d), the Alliance is focused on five operational imperatives: cognitive superiority, layered resilience, influence and power projection, cross-domain command, and integrated multi-domain defence (NATO, 2021c). These imperatives and the overarching challenges outlined in NATO2030 have led NATO to expand its operational domains from three (air, land, sea) to include cyber (NATO, 2019) and space (NATO, 2021b). However, a school of thought exists that has strongly advocated that the cognitive domain should be given equal importance and noting that it is an area where Alliance nations are particularly vulnerable (Cole & Le Guyader, 2020). Recent examples of the ever-growing threat presented by disinformation, misinformation, and propaganda, including actions taken during the COVID-19 crisis, would support such an assertion.

It is worthwhile at this point to clarify the definitions necessary to consider conflict in the cognitive domain. First, **misinformation** may be taken to mean *the* *creation and dissemination of false and/or manipulated information, without malicious intent*. Similarly, *w*eaponised misinformation, otherwise known as **disinformation**, is defined by NATO as *"the deliberate creation and dissemination of false and/or manipulated information with the intent to deceive and/or mislead"* (NATO, 2020b). The key distinguishing difference between the two is that disseminating disinformation is an act conducted with malice and forethought. The organised dissemination of disinformation is often called a **disinformation campaign**, which may also mix in selected facts so as to obscure or confuse.

On the other hand, **propaganda** disseminates information or narratives that may be factual, false, highly selected or manipulated in nature but are intended to put a country or social group in a particular light (positive or negative) or influence opinion. More succinctly, it is information with an agenda (DiResta, 2018). The more insidious modern form, **computational propaganda**, is defined as *"the assemblage of social media platforms, autonomous agents, and big data tasked with the manipulation of public opinion"* (Woolley, 2016). The employment of **computational disinformation** has meant that the costs of creating a disruptive disinformation campaign have significantly been reduced, along with a dramatic increase in effectiveness, focus and agility.

Information warfare has always been part of organized conflict (e.g., Tzu, 2018). Nevertheless, a more modern definition, drawn from Russian doctrine, defines **information warfare (or operations)** "*[as the] conscious employment of information to enable the user to achieve his political, economic, military, or other goals … [to be] conducted constantly, in peacetime, in the period of threats … and in wartime … [by] committing all available forces and software and hardware potentialities to impact the opponent's information capabilities and protect our own against similar actions by the opponent*" (Ball, 2017). Information warfare is not about armed forces, equipment or infrastructure in the physical domain but rather a part of a strategic campaign to influence, affect perceptions, and validate realities that support national power agendas (Carr, 2019). In US doctrine information warfare consists of five aspects: electronic warfare, computer network operations, PsyOps (psychological operations), military deception, and operational security (Bernal et al., 2020). More concisely, "*information warfare works to control the flow of information*" (Bernal et al., 2020; Cluzel, 2020).

One of the critical challenges of **defensive information warfare** is to recognise when an attack is occurring (Nimmo, 2016) and to act accordingly. In addition, defensive information warfare may also be used as a means of developing a narrative to inoculate one's citizens against an actual or perceived information attack (Dumitrescu, 2019).

Narratives are stories about stories (Verdon, 2019). Beliefs, history, information, disinformation, misinformation, propaganda and mythmaking are all part of the development of a **strategic narrative** defined as "*a means for political actors to construct a shared meaning of the past, present, and future of international politics to shape the behaviour of domestic and international actors. Strategic narratives are a tool for political actors to extend their influence, manage expectations, and change the discursive environment in which they operate. They are narratives about both* *states and the system itself, both about who we are and what kind of order we want. The point of strategic narratives is to influence the behaviour of others*" (O'Loughlin et al., 2013). Strategic narratives are a double-edged sword, playing a role in societies as tools for defensive information warfare (potentially based on disinformation) or as a means of increasing social resilience to cognitive conflict (Dumitrescu, 2019).

Cognitive warfare is a generalisation and extension of information warfare beyond the battlefield. It is best defined as an attack on knowledge, or more clearly, it is a means of conflict designed to impact how individuals, groups or nations think about an issue, event or situation. Ultimately, "*destabilisation and influence are the fundamental goals of cognitive warfare*" (Cluzel, 2020). The difference between information and cognitive warfare is significant (Bienvenue et al., 2018). It lies in the process of turning information into knowledge, as per the first two steps of the OODA loop (Osinga, 2005). In cognitive warfare, the weapon is information, but the domain of operation is cognitive.

Cognitive warfare is part of the larger concepts of grey zone or hybrid warfare. **Grey zone warfare** describes conflict conducted below the level of engagement that could initiate traditional military actions. *"The Gray Zone is characterised by intense political, economic, informational, and military competition more fervent in nature than normal steady-state diplomacy, yet short of conventional war"* (Votel et al., 2016), with the Cold War being the quintessential example. NATO considers **hybrid warfare** as "*a wide range of overt and covert military, paramilitary, and civilian measures … employed in a highly integrated design*" (Hoffman, 2018). This definition describes the use of mixed military-political tools of conflict used within the grey zone of operations.

2.3 Cognitive Warfare and COVID-19

NATO nations have recognised the threat of hybrid and grey zone conflict. At the 2018 Brussels Summit, heads of state noted that "*We face hybrid challenges, including disinformation campaigns and malicious cyber activities.*" This was followed up in 2019 at the London Summit with heads of state noting that NATO will be "*strengthening [its] ability to prepare for, deter, and defend against hybrid tactics that seek to undermine our security and societies*", a point reinforced once again at the 2021 NATO summit in Brussels. In the intervening years, NATO has developed a clear strategy to counter disinformation based on promulgating factual narratives. Disinformation is countered by offering a counter-narrative built on refuting, debunking, providing relevant facts and sources, and supporting the actions of a free press. Alliance nations share best practices, approaches, lessons learned, and insights among themselves and partners such as the EU and UN. This counter-disinformation strategy includes continued research in disinformation methods and grey-zone tactics. NATO S&T has played a vital role in supporting NATO efforts to understand better and counter disinformation, whether through the Science for Peace and Security Program, working with partner nations, or through the collaborative efforts of the Science and Technology Organization.

The COVID-19 pandemic is not a consequence of geostrategic conflict. Nevertheless, the pandemic has had a direct and adverse effect on all Alliance nations, particularly their medical, political, and economic systems. Exploiting this crisis, several countries and groups with partisan agendas attempted to create a watershed moment in global security by undermining Alliance cohesion and national resolve. As noted by the World Health Organization's (WHO) Director-General at the 2020 Munich Security Conference, "*We're not just fighting an epidemic; we're fighting an infodemic*" (Zarocostas, 2020).

The COVID-19 pandemic presented an information exploitation opportunity for some strategic actors as part of existing grey zone conflicts. At the same time, this has raised awareness of the enduring challenge of countering misinformation, disinformation and outright propaganda. These attacks can, and have, imposed considerable costs to the nations and has needlessly increased the death rate from COVID-19. However, in any complex crisis, such as that experienced during pandemics, information transparency and open communication builds trust and ultimately saves lives.

Over the initial phases of the COVID-19 crises, Russian disinformation campaigns focused on stories such as (EEAS, 2021): COVID-19 was developed in NATO laboratories, global elites are imposing tyranny, and the EU is disintegrating. As time went on, these attacks expanded to include those designed to increase doubt on vaccine efficacy, safety and availability, accusing countries of genocide for not using the Russian vaccine (Sputnik V), suggesting that migrants (or NATO) have been spreading the virus across Europe, as well as implying that degenerate geopolitical and commercial interests were over-riding the interests of European and North American citizens. While some of these attacks are homegrown, there has also been a trend towards amplifying attacks from secondary sources (e.g., right & left-wing groups, Iran, ISIS, China, anti-vaxxers, etc.). This approach maintains a veneer of plausible deniability and evades responsibility for accurate reporting. At the same time, these efforts have presented an inflated narrative of Russian (and to some extent Chinese) generosity and scientific, military and societal strength. This narrative was supported by logistics and medical support provided to groups with deep historical and linguistic ties to Russia or Kremlin-friendly forces (Ozawa, 2020).

In late April of 2020, Russia launched three harmonised disinformation attacks on NATO (NATO, 2020b). First, a fake letter was published, purportedly from Jen Stoltenberg, NATO's Secretary-General, to Raimundus Karoblis, Lithuania's Minister of Defence. The letter noted that NATO had decided to withdraw its multinational battlegroup from Lithuania due to rising COVID-19 infections. The letters were posted on blogs in English and Lithuanian and were later picked up by various websites and social media platforms. Fake accounts, created days before, were used to spread the disinformation further. Following this, an edited YouTube video was released suggesting that COVID-19 within the multinational battlegroup was the single agenda item for a recent NATO Defence Ministers meeting. Finally, the letter was sent directly to NATO Headquarters via a spoofed NATO e-mail address. Surprisingly, given the care taken to orchestrate this attack, the letter contained many formatting, structural and grammatical mistakes, and was recognized quickly as a fake.

Ultimately the attack was a failure and died out quickly. NATO responded rapidly to discredit the story, as did the Lithuanian Defence Ministry. Early identification (understand) and rapid reaction (engage) nullified the attack. As noted by the Secretary-General, "*They have not been successful because, first of all, it has been clearly revealed that this is a fake letter; secondly, we've seen that NATO allies remain committed, remain united, and are actually helping each other in the midst of the coronavirus crisis*" (NATO, 2020b).

Disinformation has been a constant threat to the Alliance since its inception. Nevertheless, with the increased use of asymmetric (hybrid) warfare and the attendant growth of social media, disinformation and its institutional twin of propaganda have significantly increased in volume and precision. Moreover, these efforts are part of more significant military and geopolitical trends, such as those presented by hybrid or grey zone warfare, where success is defined through operations in the cognitive sphere.

This infodemic has presented an existential crisis to Alliance nations on par with the COVID-19 pandemic itself. Throughout this crisis, NATO undertook to sustain its military and political mission, operations, and activities while maintaining operational readiness. This included using NATO strategic lift to ensure medical equipment supplies were delivered to the Alliance and partner nations. After the initial stages of the pandemic, NATO developed a refined operational plan, took stock of lessons learned, restocked medical supplies, identified funding and explored ways and means to support Alliance and national resilience. These activities created a solid narrative of an Alliance coming together to support national COVID-19 responses. Nevertheless, this positive narrative was challenged by misinformation, overt propaganda, and weaponised information campaigns, i.e., targeted disinformation.

2.4 Countering Disinformation: Engagement in the Cognitive Domain

2.4.1 A Successful Counter-Strategy

Drawing on long experience gained over its seventy-two-year history, NATO has developed a simple bifurcated strategy for countering disinformation. This strategy, built upon the classic OODA loop or Boyd Cycle (observe-orient-decide-act) (Osinga, 2005), consists of **understand** (observe-orient) and **engage** (decide-act) phases. Expanding on this (NATO, 2020b):

Understand: NATO *"Information Environmental Assessments"* systematically track, monitor and assess the information environment germane to NATO operations or policies. This observe-orient function allows NATO to identify information attacks and evaluate the impact of counter-narratives.

Engage: Once an attack is understood, a strategic approach is identified, and a fact-based counter-narrative developed. Over the longer term, this builds upon the success and lessons learned from previous disinformation counter-strategies. These counter-narratives and information products are "*fact-based, timely, transparent and coordinated*" (NATO, 2020b). This enables NATO to deliver a targeted effect within the contested cognitive domain.

The success of these efforts is greatly enhanced when coordinated across the Alliance. Consistency of message, clarity of meaning, rapid and agile reaction, and a fact-based counter-narrative are crucial to building trust and confidence. Ensuring consistency of this narrative and sharing cognitive domain awareness within the Alliance, EU, UN and other allies or partners increases the likelihood of countering a sustained and reactive disinformation campaign while building societal resilience to such attacks. Nevertheless, there are inevitably groups or populations resistant to such measures preferring conspiracy theories or biased narratives. The goal is not to convince everyone but to reach an audience receptive to a facts-based approach, thereby building trust within such communities and minimising the damage inflicted by misinformation, disinformation, or propaganda. In the end, "*Sunlight is said to be the best of disinfectants; electric light the most efficient policeman*" (Brandeis, 2014).

2.4.2 Understand: Cognitive Domain Situational Awareness (CDSA)

Situational awareness is the key to operational success, or more generally, successfully applying the instruments of national power, be they diplomatic, information, military or economic (DIME) (Worley, 2012). The famous military theorist Sun Tzu noted, "*if you know the enemy and know yourself, you need not fear the results of a hundred battles*" (Tzu, 2018). However, the modern cognitive information battleground exists in a complex web of worldwide intersecting and interconnected socio-technical networks, with state, non-state actors and empowered individuals seeking influence and attention. NATO's cognitive domain situational awareness provides it with actionable insights, helping to shape its communication response and inform the nations and partners. To counter disinformation attacks, NATO recommends that consumers of information: check the credibility of the source; assess the emotional tone; look for alternate sources, check for doctored images/video and curb their own biases (NATO, 2020b).

Regarding COVID-19, many state, non-state and agenda-driven individuals have used the crisis to spread misinformation, disinformation, and outright propaganda. These efforts are often not accidental or ill-considered. In many cases, they are deliberate actions designed to exploit modern marketing stratagems and are highly developed, employing sophisticated cyber and psychological warfare approaches. They seek strategic effects, attempting to provide a (false or misleading) narrative to create an alternative worldview for calculated impact. Such attacks are also characterised by inconsistent messaging, rapid evolutionary improvements based on message success and a shotgun approach, i.e., wide-area attacks. However, building upon modern methods, these attacks are also developing in sophistication, becoming even more targeted and effects oriented as they build upon commercial approaches, including automation.

State actors, such as Russia, employ a broad spectrum of such approaches in remarkably sophisticated fashion. Media outlets, controlled by the Russian government, such as RT (Russia Today) or Sputnik, use a broadband strategy of disinformation through a mix of news stories that blend fact with demonstrable falsehoods (NATO, 2020b). At the same time, they employ fakes sites, automated bots, and burner accounts to amplify and normalise narratives while creating momentum and uptake by other media. These later approaches may be highly targeted at specific receptive groups defined by ideology, geography, or demographics. However, the aims of such attacks are long-term and strategic, aiming to discredit NATO, the nations and Alliance values or the value of the Alliance. Sadly, in this regard, not much has changed since the cold war, except the volume, automation and targeting effectiveness of such disinformation attacks. Trust has always been challenging to build and harder to maintain.

At the start of the COVID-19 crisis, several actors took the opportunity to launch disinformation attacks on the Alliance. A surge in such attacks occurred between March and June of 2020, although the volume and nature of these attacks morphed over time. To put this in perspective, from December 2019 to April 2020, over eight thousand cases of disinformation were noted (Ozawa, 2020) by the European External Action Service's East StratCom Task Force (European External Action Service, 2021b). A review of the same database for December 2019 to May 2021 notes 11,787 instances of pro-Kremlin disinformation, with approximately 203 related to COVID or the pandemic and 1388 associated with NATO.

2.4.2.1 Russia

The use of disinformation as a Russian strategic weapon was honed during the cold war. Then again, Russia's skilful use of propaganda and disinformation goes back to Lenin (Ball, 2017) and arguably even further back to Potemkin (Montefiore, 2005). Indeed Lenin is quoted as saying, "*A lie told often enough becomes the truth*" (DiResta, 2018). Using disinformation as a strategic tool returned, with a vengeance, following the ascension of Vladimir Putin. In 2014, Russian—NATO relations were seriously strained due to the Russian annexation of Crimea. This territorial appropriation was built upon a long campaign of disinformation in the region (Ball, 2017) as part of the skilful use of hybrid warfare tactics, which included the use of "*little green men*" (Pifer, 2019) to spearhead the process. As noted at the Wales summit in 2014, by General Philip Breedlove, Supreme Allied Commander Europe (SACEUR) at the time, *"[Russia is conducting] the most amazing information warfare blitzkrieg we* *have ever seen in the history of information warfare*" (Pomerantsev, 2014). Leveraging new technologies and social media has enabled highly effective information campaigns whose implications are only beginning to be appreciated and understood (Chabuk & Jonas, 2018).

Since that time, Russian disinformation campaigns have focused on the predictable goals of undermining Alliance cohesion and stability, bring into disrepute NATO's military or political actions, spreading panic and undermining democratic processes (Monaghan, 2020; Ozawa, 2020). Strategically these efforts are designed to weaken the economic and military capability and capacity of the Alliance while creating a multi-polar world order more conducive to Russian interests as well as leaving Russia and others in a more advantageous military, economic and political position (Ozawa, 2020). In the context of COVID-19, disinformation activities that ultimately delay, weaken, prolong, or confuse will extend and amplify the pandemic and its adverse effects while provoking panic. This could eventually undermine societal bonds and degrade the geopolitical strength of Europe and North America. As noted (Pomerantsev, 2014), "*The new Russia doesn't just deal in the petty disinformation, forgeries, lies, leaks, and cyber-sabotage usually associated with information warfare. It reinvents reality, creating mass hallucinations that then translate into political action.*" Given the insignificant resource costs and minimal negative consequences of such a disinformation campaign, such a strategy is being pursued vigorously. Consequently, it is no surprise that the COVID-19 crisis became just another opportunity to be exploited.

In Spring 2021, Latvia was subject to a Russian disinformation attack, accusing Canadian Forces troops operating in Latvia of spreading COVID-19 to Latvia (NATO, 2020b). This disinformation attack consisted of a fake interview, modified screenshots from the battlegroup's FaceBook page, and an attempt to solicit a negative response from the Latvian Minister of Defence through a spoofed journalist's account.

Poland was also targeted at the same time (NATO, 2020b). In this case, a forged letter was planted, via a cyber-attack, on the Polish War Studies academy's website from Brigadier General Ryszard Parafianowicz. This letter criticised US forces in Poland, and in particular, the associated DEFENDER-Europe 20 military exercise. Once again, multiple blogs and sympathetic media outlets were used to spread the message, spoofing an e-mail from a former Polish parliamentarian to Polish media outlets.

While these targeted disinformation attacks were ongoing, in the spring of 2020, Russia continued to spread disinformation through state-influenced or controlled media channels on the nature and scope of NATO exercises in Eastern Europe in the Spring of 2020 (NATO, 2020b). NATO was accused of continuing to hold large-scale exercises with no regard for the health and safety of national populations. NATO forces were also accused of ignoring travel restrictions and safety protocols, spreading COVID-19 throughout Europe. Starting in early January of 2020, and perhaps in an even more egregious disinformation attack, pro-Kremlin media explicitly accused NATO (and the US in particular) of having engineered the COVID-19 as a biological warfare agent in secret laboratories in Latvia, Georgia, Kazakhstan, Moldova or Ukraine. The COVID-19 pandemic was portrayed as a "*biological Chernobyl*". As late as May 2020, the Russian Minister of Foreign Affairs Sergey Lavrov was quoted as saying, "*Washington's unwillingness to ensure the transparency of its military biological activities in various parts of the world raises questions about what is really going on there and what the actual goals are*" (NATO, 2020b).

Across all these targeted attacks, NATO worked with the countries involved to identify and counter the disinformation attempts before they could gain a foothold within broader media circles (NATO, 2020b). NATO also provided facts to counter the information wide-area attacks made through Russian media. In all cases, these have helped build a more fact-oriented narrative within the Alliance and Eastern Europe more specifically. However, the impact on counter-NATO narratives within Russia has perhaps been less successful (e.g. Schaffner & Flood, 2020), in no small measure due to state control or influence on major media outlets (Ball, 2017).

These Russian targeted disinformation campaigns follow a broadly similar pattern (NATO, 2020b). Such attacks employ forgeries (doctored letters, fake social media posts and bogus interviews), use burner accounts to create false personas or exploit that of a credible individual, generate explicit falsehoods and amplification of disinformation by arms-length but pro-Russian news and social media sites, engineer leaps from the originating language to English language media, and spoof e-mails to NATO, as well as governmental or media sources to provoke a reaction designed to add fuel to the controversy.

Russian disinformation has also taken on a more threatening tone (NATO, 2020b). In March of 2020, Russia offered and deployed aid supporting Italy's COVID-19 response, building upon a narrative of a responsible global actor reaching out to help Italy in its hour of need after having been abandoned by its NATO allies. Such donation diplomacy was arguably staged as a pure propaganda exercise (Weiss, 2020). Still, it must also be noted that NATO and the EU were slow in responding to Italian requests for aid, thus creating an opportunity to be exploited. As such, the Italian government initially welcomed this assistance as a good-faith gesture. Soon, however, questions were being asked in mainstream Italian media about the usefulness of such aid, initial deployment near a NATO airbase far from the centre of the outbreak and the potential embedding of GRU agents within such support (NATO, 2020b; Braw, 2020). Following a strategy of "*the best defence is a good offence*", senior Russian officials condemned the reporting, accusing the Italian media of spreading false news and disinformation (NATO, 2020b). The original author of the critical reporting was threatened on Facebook by the Russian Ministry of Defence Spokesperson, Igor Konashenkov. Soon after the Spokeswoman for the Ministry of Foreign Affairs of the Russian Federation, Maria Zakharova, accused multiple US media outlets of spreading disinformation on Russian COVID mortality rates, threatening action unless the articles were retracted. In the context of these events, NATO leadership reiterated that the fourth estate has a responsibility to check and assess information on behalf of free societies. By taking an open and fact-based approach to disinformation NATO demonstrated its willingness to put itself under a microscope. Action, ultimately, it is hoped, speaks louder than words.

Disinformation is a strategic priority within Russia. Such disinformation has two strategic objectives. First, Russia's return as a great power, with a commensurate return of Russian influence in Eastern Europe and the collapse of NATO as an effective organisation (Ball, 2017). Second, to preserve the Putin regime by protecting it from the contagion of western-oriented revolutions (Ball, 2017). To achieve these objectives, a false narrative is presented where the US is a demonic military-industrial force for global tyranny, EU members are at each other's throats, Eastern Europe is a hotbed of ethnic tension, NATO is threatening an imminent attack on Mother Russia, Ukraine (with a population approximately 28% of Russia) presents a credible military threat, Alliance democracies have failed to tackle the COVID crisis and western vaccines are dangerous and produced by corrupt manufacturers. This picture is contrasted with a militarily and politically strong Russia, prepared to counter NATO aggression, a friend always willing to lend a hand during a crisis, leading the world in developing COVID-19 vaccines, and providing an effective national response to the situation. Such a Potemkin narrative has real power based on a blend of carefully selected facts, half-truths, cultural bias, outright falsehoods and wish fulfilment. Note that while there is an overarching strategy to use disinformation, there is also an element of opportunism. "*Putin is not a "cool genius … His system is 'adhocracy,' in which lackeys do not receive direct instructions but instead rely on hints and guesses to determine what will please the boss*" (Galeotti, 2019; Lipman, 2019). Putin is an opportunist without a master plan, but with a commitment to restoring Russian greatness. Russian actions during the COVID-19 crisis are simply the most recent example of this ad hockery.

Beyond the COVID-19 crisis, Russia views disinformation campaigns as a fundamental aspect of ongoing grey-zone or hybrid warfare. In this world, "*truth is negotiable … a popular uprising a fascist coup and … [support to] a Syrian dictator becomes an anti-terrorist operation*" (Ball, 2017). The Russian Chief of the General Staff, Valery Gerasimov, is quoted as saying, "*interstate confrontation's foundation has come to consist of non-military measures including political, economic, and informational … [Total conflict] has spread to all spheres of activity in modern society, diplomatic, scientific, sport, and cultural and, in fact, has become total in scope*" (Ozawa, 2020). This comment contains the echoes of a cold-war long thought confined to history.

2.4.2.2 China

As with Russia (a diminished global power), China (a growing global power) has exploited the geostrategic opportunities presented by the global crisis. For China, disinformation and dissembling appear to serve a larger strategic geopolitical purpose. Having been (most likely) the source of COVID-19, China appears to have emerged more quickly from the pandemic. As such, China's information campaign seems to be based around a more assertive diplomatic stance (i.e., "*Wolf Warrior*" diplomacy), exploitation of a distracted world to attain or advance territorial goals (e.g., the India-China border dispute, de facto integration of Hong Kong, the South China Sea, Taiwan, and the Senkaku (Diayu) Islands) and to restore China to its rightful place in the international community after a century of humiliation (while protecting the image of the Chinese Communist Party (CPP)). The associated broadband information campaigns seek to shield China from criticism or more profound exploration of their management of the initial stages of the pandemic. At the same time, it portrays China as a rising global leader with a demonstrably superior governance and developmental model.

The Chinese Communist Party (CPP) has long used propaganda and disinformation to shape domestic and international perspectives. "*[The CPP has] long prioritised disinformation to advance its domestic monopoly on power, claims to global leadership, and irredentist aspirations. It has built a sprawling infrastructure for manipulating information and disseminating its preferred narratives both at home and abroad*" (Diresta et al., 2020).

During the pandemic, China's disinformation campaign has sought to obfuscate investigations around the source of the COVID-19 virus, generally accepted to have emerged originally in Wuhan (WHO, 2021). Without evidence, both the Global Times (state-owned) and the Foreign Ministry (Zhao Lijian) have claimed that the virus originated outside of China (either Italy or the US). They have also worked hard to discredit the response to COVID-19 by Alliance nations and partners, claiming superior Chinese vaccines and a more effective societal response (NATO, 2020b). NATO has been singled out as having provided ineffective support to the nations and partners (NATO, 2020b). Like Russia, China has taken the opportunity to conduct highly effective propaganda exercises, donating health care supplies early in the pandemic and vaccines (Sinovac) later on to several countries. This vaccine diplomacy has supported a soft-power narrative of a country with superior political, technological, and economic systems. However, China's vaccine diplomacy has also come with significant strings attached (Weiss, 2020).

2.4.2.3 Non-State Actors

Nation-states are not the only actors seeking to exploit the pandemic. Non-state actors, such as terrorist groups, criminal groups, commercial entities, and hyper-empowered individuals, have all sought to take advantage of the pandemic (NATO, 2020b). Examples may be found in the claims by ISIS that COVID-19 presents an opportunity to conduct new attacks, especially as these will make them immune to COVID-19. Building on a narrative that COVID-19 is God's punishment for unbelievers, they have sought to increase recruitment and supporters for their terrorist activities. Other examples of such a marriage between religion and the pandemic may also be found within Alliance countries, with fringe groups seeking to exploit the pandemic for political or social purposes. Denial of the pandemic itself and efficacy of basic public safety measures has been fuelled by an infodemic of misinformation, conspiracy theories and disinformation. Often this disinformation has clear political or commercial motivations.

An under-appreciated actor in this context is that of the anarchist, *information graffiti-artist* or *internet troll*. Such hyper-empowered individuals disrupt the information space by adding disinformation for the sake of causing damage and chaos, or in some cases, a strong belief in a *cause*. They may also play a critical role in amplifying disinformation obtained from other sources or (in the pay of national actors) deliberately shaping, manipulating and obfuscating a factual narrative. In this latter role, China has been exposed as using such paid influencers to downplay the initial severity of the COVID crisis (Zhong et al., 2020).

2.4.3 Engage: Counter-Narratives

NATO's approach, honed over seventy-two years of information conflict, is based on a traditional approach of highlighting concrete actions, verifiable facts and credible communication. This approach relies on clear and consistent narratives passed through conventional mass media, public engagements and social media. Such an approach is based on the Alliance's ideals and core values. Trust, once broken, is hard to regain, so clarity and consistency are central aspects of this approach. For example (NATO, 2020b), in support of the Alliance and its partners, "*NATO and Allied armed forces have played a key role in supporting civilian efforts to fight COVID-19, with some 350 flights delivering hundreds of tons of critical supplies, the construction of almost 100 field hospitals, and almost half a million troops across the Alliance securing borders, transporting patients and helping with testing [and vaccine delivery].*" It has also built over twenty-five field hospitals, provided over twenty-five thousand beds to the Allied nations, and supports sharing medical supplies and logistics through the NATO Euro-Atlantic Disaster Response Coordination Centre (EADRCC) (Ozawa, 2020).

Responding to specific attacks, NATO provides widely available refutations, briefings, talking points, statements, and corrections. In response to the COVID-19 infodemic, NATO intensified its communications across a wide range of platforms and media, engaged in an increasing number of online engagements with a broad range of audiences. Also, it expanded outreach within Russia through various media products. However, this has not been enough, and NATO has reached out to academia, industry, NGOs, think tanks, and media watch organisations to encourage discussion. This can reduce the susceptibility to disinformation and support a more resilient society. An engaged, independent, and empowered fourth estate is a necessary part of this resilience. Only by constant engagement with old and new media across the Alliance, partners and where possible, social groups/countries that are the sources can one hope to ensure that disinformation is successfully countered.

The response to disinformation has been coordinate through the Strategic Communication Task Force, led by the NATO Public Diplomacy Division (PPD). An example of NATO's approach may be found in its straightforward debunking of Russian myths on NATO and COVID-19 (NATO, 2020a). This strategy faces the recurring issues raised in Russian disinformation efforts and attempts to answer the accusations in a forward-leaning fashion.

In the end, NATO has faced the COVID pandemic in two ways (Ozawa, 2020). The first is by leveraging the power of the military Alliance to support a global medical emergency, and the second by addressing the infodemic head-on. The efforts of the S&T community have successfully been brought to bear on both issues.

No single strategy will work all the time, nor will counter-narrative strategies be universally successful, as exemplified by the continuing existence of "*flat earth*" societies despite irrefutable evidence to the contrary. NATO's engagement strategy, therefore, relies on proactive and rapid communication, highlighting flawed reasoning (debunking), identifying familiar sources and methods of disinformation, countering disinformation on the platform or via the media on which it is found, adapting to the evolution of such disinformation, being selective in which attacks to counter and consistent in overall messaging. Not all disinformation can be countered, if for no other reason than the sheer volume of such occurrences, so selecting those attacks that can do the most harm or are starting to gain significant attention is a critical step. Earning the trust of Alliance societies through consistent communication and provision of information will help slow or marginalise the remainder.

2.4.4 Lessons Learned

COVID-19 presented a medical emergency the likes of which Europe and North American had not faced since the Spanish Flu outbreak early in the twentieth century. Early on missteps were made by NATO and the nations in dealing with the evolving crisis, but the Alliance adapted quickly. Rich lessons have been learned, and mistakes quickly corrected (Ozawa, 2020).

First among these lessons is that timing is essential, as are symbolic gestures. Responding promptly with concrete actions is the best way to disable an information attack and build trust within the Alliance.

Second, organisational relationships are essential and need to be established (and maintained) in advance. Crises such as COVID-19 are "*come as you are events.*" Leveraging existing relationships with the EU, industry, academia, NGO, UN, WHO, and media are key to a rapid and effective response. Open, transparent and ongoing communication at all levels supports these relationships and allows for the rapid creation of "trust networks" (Little, 2010; Tilly, 2005) essential to a prompt, agile and comprehensive multilateral whole-of-government response to a crisis.

Third, COVID-19 has demonstrated the value of collective action. Nevertheless, there is now an opportunity to take the lessons learned during this crisis into account as NATO charts a new path forward as part of the NATO2030 initiative. As noted by the NATO Secretary-General Jens Stoltenberg, "*NATO is continuing to adapt to keep us safe in this decade and beyond. The NATO2030 initiative is about making sure our Alliance remains ready today to face tomorrow's challenges*" (NATO, 2021a; Maizière et al., 2020).

Fourth, NATO needs to better adapt to hybrid, grey zone and cognitive warfare. The battleground of today and the future includes the sixth operational domain, the cognitive or, more generally, the human, domain. Russian and Chinese approaches to disinformation have been successful to some degree, impacting the lives of citizens in the Alliance and with allies. In a manner consistent with the values and ethics that underlie the Alliance, mastering this domain will be essential to maintaining the Alliance's effectiveness and reducing the risk of strategic errors with potentially disastrous repercussions.

Finally, the current approaches to disinformation rely on tested methods of countering lies with facts. However, real-world examples of social behaviour and groups have demonstrated facts are often ineffective against communities with strong biases. Individuals or groups strive to maintain a sense of competence by actively resisting information contrary to their beliefs (Lishego, 2019). This has been exacerbated by modern media as it allows for a selection of news, opinions and facts tailored to our interests and beliefs. As a result, there is considerable opportunity for more modern scientific and technical methods to bear on the problems of measuring the impact of disinformation, narrative clashes and understanding cognitive biases.

2.5 S&T Support

COVID-19 presented a unique test of the NATO S&T community, with the Secretary-General and North Atlantic Council calling for the whole intellectual and scientific weight of the Alliance to be brought to bear on the problems associated with COVID. As noted by the Chairman of the NATO Military Committee Air Chief Marshal Sir Stuart Peach, Military Committee in Chiefs of Defence Session, 14 May 2020, "*NATO has tapped into its very important resource of defence scientists—the largest such network in the world—to support the COVID-19 response.*" Many in NATO's network of over 6,000 scientists and engineers rose to the challenge of supporting efforts to combat the effects of COVID-19 (NATO, 2020c).

Throughout the pandemic, the NATO S&T Organisation (STO), along with the Science for Peace and Security Program, turned their wide-ranging network of scientists, research partners and institutions to explore innovative ways to support NATO as it worked with nations to address the pandemic. This work focused on addressing the most pressing problems associated with the pandemic. Problem areas covered a broad range of activities such as understanding disinformation methods, countermeasures and impact, sustaining the health of NATO forces during relief operations, conducting analysis of current pandemic development and planning for future pandemics, improving the use of technology for pandemic relief operations, participating in lessons learned activities on NATO's response and considering moral and ethical injury during pandemic relief operations.

One study undertaken by the STO Systems Analysis and Studies (SAS) panel investigated possible post-Covid-19 futures, exploiting the network's unique abilities to help address the post-COVID-19 geopolitical, military and economic strategic landscape. For this purpose, a quick reaction specialist team was established to identify potential futures for the aftermath of COVID-19 and the impacts on defence. The scientific experts from Allied and Partner Nations that formed this specialist team worked with the following NATO Bodies: Allied Command Transformation (ACT), Joint Analysis and Lessons Learned Centre (JALLC), and the NATO Communications and Information Agency (NCIA). Together, they developed possible futures for the coming six years, focused on the post-COVID-19 environment, and analysed the potential military strategic and operational impacts, including cognitive warfare.

Another activity of note was conducted by the STO Human Factors and Medicine Panel (HFM-336), which leveraged their research in disinformation to advise NATO's COVID-19 response task forces. The team also hosted a series of panel sessions to share best practices and insights on the nature, countermeasures, and best practices with the nations and senior leaders. A video was even produced, "How is NATO Responding to Disinformation on COVID-19?" to provide leadership with a better understanding of misinformation about the pandemic and how to counter it. These efforts improved the Alliance's and national abilities to monitor and counter disinformation, which otherwise would have exploited public uncertainty over health undermining the credibility and effectiveness of national actions.

Research in this area is expected to continue, if for no other reason than to sharpen NATO's ability to understand and counter disinformation activities. NATO's disinformation strategy is mainly based on a traditional focus on facts and clarity of message. This strategy assumes a rational actor's response to facts and evidence. However, it is clear from more recent research, and experiences with disinformation campaigns that falsehoods travel faster than facts (Vosoughi et al., 2018) and belief (coupled with a sense of belonging to a social group) is more potent than facts (Fisher, 2021). More critically, "*Today's emergent technologies are enhancing those longstanding capabilities, enabling greater velocity and virality, and offering access to new audiences and ways of spreading information*" (Diresta et al., 2020). Thus, a better understanding of how modern disinformation spreads and the impact of new technologies will need to be developed if NATO is to counter future disinformation campaigns and hostile strategic narratives successfully.

2.6 The Impact of Emerging and Disruptive Technologies

Emerging and disruptive technologies (EDTs) will have an outsized effect on developing new methods for countering disinformation operations and strategic computational propaganda in the context of strategic cognitive warfare. The problems associated with misinformation and disinformation will increase dramatically in the future as more sophisticated methods are employed (Bradshaw & Howard, 2019). These will draw from modern psychological and social research, tested and perfected in the digital bazaar. Precision information warfare will emerge from these new embryonic technologies.

Within NATO, there are seven canonical EDTs (Reding & Eaton, 2020). These EDTs are Data (Big Data and Advanced Analytics), Autonomy (Autonomous Systems and Robotics), AI (Artificial Intelligence), Quantum (Next-Generation Quantum Technologies), Space (Space Technologies), Hypersonics (Hypersonic Weapon Systems) and Biotech (Biotechnologies and Human Enhancement Technologies). In addition, Materials (Novel Materials and Manufacturing) research has been added to this for considerations within the S&T Organization. The following sub-sections outline some of the foreseen impacts of EDTs on the future cognitive arena based on NATO technology watch and foresight activities (Reding & Eaton, 2020).

2.6.1 Data

Leveraging new computational and communication modes, including AI, will significantly expand the weaponisation, precision targeting and sophistication of disinformation attacks. Much like the current state of cyber, increasing sophisticated tools will be widely available to "*create and manipulate information at scale*" (Deloitte, 2019). However, these tools will be tuned for precision effect based on the analysis, modelling and simulation built on large scale analysis of social media derived data. This trend can be seen in operation today where social media information and global positioning information (e.g., proximity to religious or social institutions) can be fused to precisely target susceptible individuals. More accurate feedback on the effects of a disinformation campaign will be possible due to large volumes of target data. This increased sophistication will support the creation of increasingly effective disinformation and counter-disinformation strategies through faster and more comprehensive feedback loops. It will also provide a foundation for a deeper understanding of what constitutes effective disinformation and what defines its uptake. This data-driven approach also provides a path for the potential tactical use of precisely targeted disinformation attacks on forces and personnel in an operational theatre or their families (Fleitz, 2014).

2.6.2 Artificial Intelligence (AI)

The development of AI-enabled bots (already widely in use) simulating human responses and creating disinformation or information blizzards will become more advanced and difficult to characterise as artificial. The broader use of deep fakes (voice, video, imagery, etc.) will not only enable the creation of near-perfect disinformation but, in aggregate, will undermine faith in all forms of communication. AI-enabled tools for disinformation creation will also allow a deeper understanding of cognitive targets' social, linguistic, and cultural aspects. This AI approach will ultimately increase the effectiveness of disinformation campaigns by creating more resonant and tailored strategic narratives.

The targets of such disinformation campaigns will also change. Today disinformation seeks to change the human cognitive landscape. In the future, the victims may be AI (Vinci, 2020), either directly (via cyber) or using computational propaganda and disinformation as a means of polluting the data used to train or evolve AI systems. (Machine) cognitive warfare may indeed turn out to be more impactful and insidious than that directed at the human cognitive landscape, as its influence may be more challenging to detect.

2.6.3 Quantum

Quantum computing has the potential to revolutionise optimisation, AI, modelling and simulation. As a result, increasingly sophisticated predictive models and simulations will be available to model complex social systems with the goal of better understanding disinformation propagation and uptake. Therefore, such tools will be accessible to create increasingly more effective disinformation campaigns, including those building upon a more strategic effects-based approach to targeting.

2.6.4 Space

More widely available space derived data (e.g., SIGINT, GEOINT, IMINT, etc.) will provide an additional means to assess the impact of disinformation campaigns. For example, tracking the movement of refugees or migrants targeted by destabilising information campaigns would support the feedback necessary to calibrate even more effective campaigns encouraging such movement. More insidiously, this could be used to create a separate narrative for a second group to promote violence or an extreme social reaction.

The wide availability of satellite-based internet will open up social media to those currently with limited or restricted access. This conduit will allow another vector for disinformation flow and cognitive influence.

2.7 Conclusion

COVID-19 has been a tragedy for the world. Compounding this has been the scourge of misinformation and its weaponised form, disinformation. The COVID-19 crisis has brought together two of the most significant trends in emerging and disruptive military technologies: biotechnologies that underlie the mRNA vaccines that may ultimately reign in the virus and cognitive warfare with disinformation as the weapon of choice.

Disinformation campaigns are hardly new and have long been a staple of geopolitical conflict. Nevertheless, the precision, breadth, volume, automation and audacity of such attacks are unprecedented. Exploiting the situation, some states, social groups and individuals with malice and forethought sought to conduct cognitive warfare on entire populations, using the opportunities presented by COVID-19 to seek political or social advantage. These have had, and will continue to have, severe consequences for Alliance nations and the health of its citizens. Ultimately, as we have discussed in this chapter, the NATO alliance has faced these issues through the active use of S&T and a focus on facts, concrete actions, and trust-building. Nevertheless, it is a well understood psychological phenomenon that facts do not always win an argument. While arguably a successful strategy, the lingering effects of these attacks suggest that the struggle is far from over. In the end, whether it is at the strategic or tactical level, conflict is a decidedly human activity. The most effective way to overcome an opponent is to impact their thoughts and beliefs, thereby turning them against themselves. Continued research into disinformation and its impact on societies will help develop new battle plans to counter these attacks. Further, the success and increasing development of more sophisticated attacks suggest that cognitive warfare will become even more important for the Alliance as potentially the sixth operational domain for NATO.

References

Ball, D. Y. (2017). *Protecting falsehoods with a bodyguard of lies: Putin's use of information warfare* (Research Paper 136). NATO Defence College. https://www.ndc.nato.int/news/news.php?icode=1017

Bernal, A., Carter, C., Singh, I., Cao, K., & Madreperla, O. (2020). *Cognitive warfare: An attack on truth and thought*. NATO / John Hopkins University. https://www.innovationhub-act.org/sites/default/files/2021-03/Cognitive%20Warfare.pdf

Bienvenue, E., Rogers, Z., & Troath, S. (2018, September 19). Cognitive warfare. Government of Australia—Australian Army. *The Cove*. https://cove.army.gov.au/article/cognitive-warfare

Bradshaw, S., & Howard, P. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation* (Working Paper Computational Propaganda Research Project). University of Oxford. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf

Brandeis, L. D. (2014). *Other people's money and how the bankers use it* (1st ed.). CreateSpace Independent Publishing Platform.

Braw, E. (2020, March 30). Beware of bad Samaritans. Voice. *Foreign Policy*. https://foreignpolicy.com/2020/03/30/russia-china-coronavirus-geopolitics/

Carr, H. (2019, April). Arctic sovereignty and information warfare—The new battlespace in the North. *The Three Swords Magazine—NATO Joint Warfare Center*.

Chabuk, T., & Jonas, A. (2018, August 28). Understanding Russian information operations. *SIGNAL Magazine*. https://www.afcea.org/content/understanding-russian-information-operations

Cluzel, F. (2020). *Cognitive warfare*. NATO Allied Command Transformation. https://www.innovationhub-act.org/sites/default/files/2021-01/20210122_CW%20Final.pdf

Cole, A., & Le Guyader, H. (2020). NATO's sixth domain of operation. *NATO Innovation Hub*. https://www.innovationhub-act.org/sites/default/files/2021-01/ENG%20ultimate.pdf

Deloitte. (2019). Future of risk in the digital era. *Deloitte*. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-rfa-future-of-risk-in-the-digital-era-report.pdf

DiResta, R. (2018, October 9). Computational Propaganda. *The Yale Review*. https://yalereview.yale.edu/computational-propaganda

Diresta, R., Miller, C., Molter, V., Pomfret, J., & Tiffert, G. (2020). *Telling China's story: The Chinese Communist Party's campaign to shape global narratives*. White Paper. Stanford Internet Observatory. https://cyber.fsi.stanford.edu/publication/telling-chinas-story

Dumitrescu, L. (2019). *The role of strategic narratives in information warfare*. https://www.researchgate.net/publication/335383470_THE_ROLE_OF_STRATEGIC_NARRATIVES_IN_INFORMATION_WARFARE

Elkins, L. (2019, November). The 6th warfighting domain. *Over the Horizon Multi-Domain Operations and Strategy*.

EEAS (European External Action Service). (2021a). *EEAS special report update: Short assessment of narratives and disinformation around the COVID-19 pandemic* (Update December 2020—April 2021). European External Action Service. https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic-update-december-2020-april-2021/

European External Action Service. (2021b, May 9). *EUvsDisinfo Disinfo Database*. EU vs DISINFORMATION. https://euvsdisinfo.eu/disinformation-cases/

Fisher, M. (2021, May 7). 'Belonging is stronger than facts': The age of misinformation. *The New York Times*, sec. World. https://www.nytimes.com/2021/05/07/world/asia/misinformation-disinformation-fake-news.html

Fleitz, F. (2014, October 9). The ISIS threat to military families. *Center for Security Policy* (blog). https://centerforsecuritypolicy.org/the-isis-threat-to-military-families/

Galeotti, M. (2019). *We need to talk about Putin: How the West gets him wrong*. Ebury Press.

Galvão, J. (2020). COVID-19: The deadly threat of misinformation. *The Lancet Infectious Diseases, 21*(5). https://doi.org/10.1016/S1473-3099(20)30721-0

Hoffman, F. G. (2018). Examining complex forms of conflict: Gray zone and hybrid challenges. *PRISM | National Defense University, 7*(4). https://cco.ndu.edu/News/Article/1680696/examining-complex-forms-of-conflict-gray-zone-and-hybrid-challenges/

Lipman, M. (2019, November 1). We need to talk about Putin. *Foreign Affairs*. https://www.foreignaffairs.com/reviews/capsule-review/2019-10-15/we-need-talk-about-putin

Lishego, D. (2019, April 10). Dietrich Dorner & The logic of failure. *Medium*. https://medium.com/@dlishego/dietrich-dorner-the-logic-of-failure-9cda6a9360cc

Little, D. (2010, April 9). Understanding society: Trust networks. *Understanding Society*. https://understandingsociety.blogspot.com/2010/04/trust-networks.html

Maizière, T., Wess Mitchell, A., Bew, J., Bossenmaier, G., Dalgaard-Nielsen, A., Dassù, M., Fotyga, A., Ildem, T., & Verhagen, H. (2020). *NATO2030—reflection group—Final Report.Pdf*. NATO.

Monaghan, A. (2020). *Russian grand strategy and the COVID crisis* (Policy Brief 22). NATO Defence College. https://www.ndc.nato.int/news/news.php?icode=1507

Montefiore, S. S. (2005). *Potemkin: Catherine the great's imperial partner*. Vintage.

NATO. (2019, February 12). NATO review—NATO's role in Cyberspace. *NATO Review*. https://www.nato.int/docu/review/articles/2019/02/12/natos-role-in-cyberspace/index.html

NATO. (2020a, April). *Russia's top five myths about NATO & COVID-19*. https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/2004-Factsheet-Russia-Myths-COVID-19_en.pdf

NATO. (2020b, July 17). *NATO's approach to countering disinformation*. NATO. http://www.nato.int/cps/en/natohq/177273.htm

NATO. (2020c, September 17). *Fighting COVID-19 with science*. NATO. http://www.nato.int/cps/en/natohq/news_177729.htm

NATO. (2020d, September 24). *Strategic concepts*. NATO. http://www.nato.int/cps/en/natohq/topics_56626.htm

NATO. (2021a). *NATO2030*. https://www.nato.int/nato2030/index.html

NATO. (2021b, April 22). *NATO's approach to space*. NATO. http://www.nato.int/cps/en/natohq/topics_175419.htm

NATO, ACT (Allied Command Transformation). (2021c, March 15). *NATO military leaders discuss warfare development agenda: NATO's Act*. https://www.act.nato.int/articles/nato-military-leaders-discuss-warfare-development-agenda

Nimmo, B. (2016). Identifying disinformation: An ABC. *Institute for European Studies, 1*(2016), 8.

O'Loughlin, B., Miskimmon, A., & Roselle, L. (2013). *Strategic narratives: Communication power and the new world order*. https://doi.org/10.4324/9781315871264

Osinga, F. (2005). *Science, strategy and war: The strategic theory of John Boyd*. Eburon Academic Publishers.

Ozawa, M. (2020). *COVID-19 NATO in the age of pandemics: NATO and Russia in the time of corona: Countering disinformation and supporting allies* (NDC Research Paper 9). NATO Defense College. https://www.ndc.nato.int/download/downloads.php?icode=642

Pifer, S. (2019, March 18). Five years after crimea's illegal annexation, the issue is no closer to resolution. *Brookings* (blog). https://www.brookings.edu/blog/order-from-chaos/2019/03/18/five-years-after-crimeas-illegal-annexation-the-issue-is-no-closer-to-resolution/

Pomerantsev, P. (2014, September 9). How Vladimir Putin is revolutionising information warfare. *The Atlantic*. https://www.theatlantic.com/international/archive/2014/09/russia-putin-revolutionizing-information-warfare/379880/

Reding, D., & Eaton, J. (2020). *Science and technology trends 2020–2040: Exploring the S&T edge*. NATO Science and Technology Organisation. https://www.sto.nato.int/pages/tech-trends.aspx

Schaffner, T., & Flood, A. (2020, May 5). *Poll: Majority of young Russians distrust NATO, don't consider Russia a European country | Russia matters*. Russia Matters, Harvard Kennedy School, Belfer Center for Science and International Affairs. https://www.russiamatters.org/blog/poll-majority-young-russians-distrust-nato-dont-consider-russia-european-country

Tilly, C. (2005). *Trust and rule*. Cambridge University Press.

Tzu, S. (2018). *The art of war* (1st ed.). Samaira Book Publishers.

Verdon, T. (2019, April). What is a/the narrative: A story about stories. *The Three Swords Magazine—NATO Joint Warfare Center*.

Vinci, A. (2020, October 7). *The coming revolution in intelligence affairs*. https://www.foreignaffairs.com/articles/north-america/2020-08-31/coming-revolution-intelligence-affairs

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Votel, J. L., Cleveland, C. T., Connett, C. T., & Irwin, W. (2016). Unconventional warfare in the gray zone. *Joint Force Quarterly*, no. 1st Quarter: 9.

Weiss, M. (2020, June 29). The Sinister fiction of Russian aid to the COVID-plagued world. *The Daily Beast*, sec. world. https://www.thedailybeast.com/the-sinister-fiction-of-russian-aid-to-the-covid-plagued-world

WHO (World Health Organisation). (2021). *WHO-convened global study of origins of SARS-CoV-2: China Part*. Joint Report. World Health Organisation.

Woolley, S. C. (2016). Political communication, computational propaganda, and autonomous agents—Introduction. *International Journal of Communication, 10*, 9.

Worley, D. (2012). *Orchestrating the instruments of power: A critical examination of the U.S. National Security System*. https://www.researchgate.net/publication/298705751_Orchestrating_the_instruments_of_power_A_critical_examination_of_the_US_national_security_system

Zarocostas, J. (2020). How to fight an infodemic. *The Lancet, 395*(10225), 676. https://doi.org/10.1016/S0140-6736(20)30461-X

Zhong, R., Mozur, P., & Kao, J. (2020, December 19). Leaked documents show how China's army of paid internet trolls helped censor the coronavirus. *ProPublica & New York Times*. https://www.propublica.org/article/leaked-documents-show-how-chinas-army-of-paid-internet-trolls-helped-censor-the-coronavirus?token=6ztk8omKweU8rIcm8bBlBXRV8xDBAzfX

Chapter 3
Developing Approaches to Detect and Mitigate COVID-19 Misinfodemic in Social Networks for Proactive Policymaking

Nitin Agarwal, Esther Mead, Billy Spann, and Kate Donoven

Abstract Throughout the COVID-19 pandemic, people have grown more reliant on social media for obtaining news, information, and entertainment. However, the information environment has become a breeding ground for disinformation tactics. Formal recommendations from medical experts are becoming muffled by the avalanche of toxic content and social media echo chambers are being created in hopes that users only consume stories that support certain beliefs. Despite the advantages of utilizing online social networks (OSNs), a consensus is emerging suggesting the presence of an ever-growing population of malicious actors who utilize these networks to spread misinformation and harm others. These actors are using advanced techniques and are engaging on multiple platforms to propagate their disinformation campaigns. As such, researchers have had to evolve their methods to detect disinformation. In this chapter, we present novel multimethod socio-computational approaches to analyze disinformation content and actors on OSNs during the initial months after COVID-19 was made public. These techniques are presented as case studies in narrative analysis of COVID-19 misinformation themes on blogs, identifying anti-lockdown protestor coordination through connective action on Twitter, analysis of hate speech and divisive discourse on YouTube through toxicity analysis, and modeling of misinformation contagion using an epidemiological approach. We end the chapter by presenting a COVID-19 misinformation tracker tool developed in collaboration with the Arkansas Office of the Attorney General. Our results offer policymakers valuable data to make informed decisions about the information environment and derive appropriate and timely countermeasures to combat insidious forms of cyber threats. Our efforts demonstrate that when researchers coordinate with policymakers it can make a difference, especially when that coordination remains an ongoing process.

N. Agarwal (✉) · E. Mead · B. Spann
Collaboratorium for Social Media and Online Behavioral Studies, University of Arkansas—Little Rock, Little Rock, AR, USA
e-mail: nxagarwal@ualr.edu

K. Donoven
Arkansas Office of the Attorney General, Little Rock, AR, USA
e-mail: Kate.Donoven@ArkansasAG.gov

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_3

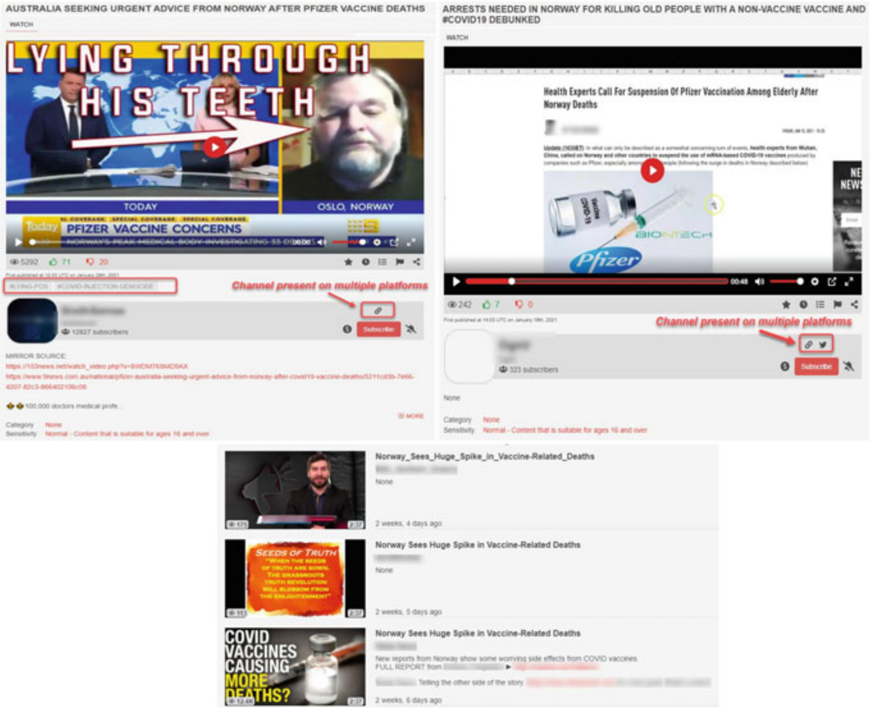
3.1 Introduction

Social media is characterized as a powerful online interaction and information exchange medium. However, it has given rise to new forms of deviant behaviors such as spreading fake news, rumors, misinformation, disinformation, and conducting propaganda and influence campaigns. Due to afforded anonymity and perceived diminished personal risk of connecting and acting online, deviant groups are becoming increasingly common. Online deviant groups have grown in parallel with online social networks, whether it is black hat hackers using Twitter to recruit and arm attackers, announce operational details, coordinate cyber-attacks (Calabresi, 2017), and post instructional or recruitment videos on YouTube targeting certain demographics; or state/non-state actors' and extremist groups' (such as ISIS) savvy use of social communication platforms to conduct phishing operations, such as viral retweeting of messages containing harmful URLs leading to malware (Al-khateeb et al., 2016). These campaigns use a variety of tactics, techniques, and strategies to further their agenda. Some of these campaigns exploit deep-rooted biases along racial, ethnic and political lines (leveraging pandemic disparity), to erode trust in scientific and democratic institutions (invoking vaccination fears), to stoke anger, anxiety, and chaos, and ultimately polarize an already divided society. People increasingly obtain their news from social media rather than from mainstream media (Barthel & Mitchell, 2017), and with the world adjusting to the current pandemic and officials struggling with the misinfodemic, it is important now, more than ever, to closely monitor social media platforms to identify misinformation and disinformation campaigns and stem their impact on society.

The use of misinformation, disinformation, and propaganda is not new. Grant (2004) discussed how Greek and Roman historians often deliberately used misinformation and disinformation in their writings. Some authors date the start of the use of misinformation back to after Johannes Gutenberg invented the printing press in 1439 (Soll, 2016).[1] Numerous research groups, such as EUvsDiSiNFO[2] have been tracking Russia's use of disinformation over the years. There are also numerous researchers that study the effects of misinformation in the political realm, such as The Harvard Kennedy School (Burton & Koehorst, 2020). The use of disinformation has also had a great impact on the Indo-Pacific Countries (INDOPAC), which is highly researched (Cha et al., 2020). The role of media in the spread of misinformation is of great concern, and is the focus of numerous studies, especially in terms of strategies to increase media literacy (Mihailidis & Viotty, 2017), and the effects of media in times of health crises such as COVID-19 (Bao et al., 2020). Various mainstream media outlets have highlighted the problem of misinformation

[1] Blackbird.AI provides Disinformation Defense and Response to national security and enterprise customers. https://www.blackbird.ai/history-of-misinformation/.

[2] EUvsDisinfo is the flagship project of the European External Action Service's East StratCom Task Force established in 2015 to better forecast, address, and respond to the Russian Federation's ongoing disinformation campaigns affecting the European Union. https://euvsdisinfo.eu/.



AUSTRALIA SEEKING URGENT ADVICE FROM NORWAY AFTER PFIZER VACCINE DEATHS

WATCH

5292 71 20

Channel present on multiple platforms

12827 subscribers

MIRROR SOURCE:

100,000 doctors medical profe...

Category None
Sensitivity Normal - Content that is suitable for ages 16 and over

ARRESTS NEEDED IN NORWAY FOR KILLING OLD PEOPLE WITH A NON-VACCINE VACCINE AND #COVID19 DEBUNKED

WATCH

Health Experts Call For Suspension Of Pfizer Vaccination Among Elderly After Norway Deaths

242 7 0

Channel present on multiple platforms

323 subscribers

None

Category None
Sensitivity Normal - Content that is suitable for ages 16 and over

Norway_Sees_Huge_Spike_In_Vaccine-Related_Deaths
None
2 weeks, 4 days ago

Norway Sees Huge Spike in Vaccine-Related Deaths
None
2 weeks, 5 days ago

Norway Sees Huge Spike in Vaccine-Related Deaths
New reports from Norway show some worrying side effects from COVID vaccines.
FULL REPORT from
Telling the other side of the story
2 weeks, 6 days ago

Fig. 3.1 Evidence of misinformation about COVID-19 vaccines killing people in Norway. Misinformation videos are pushed on multiple platforms as noted in the images above. Further, note that all the videos are of equal run time, i.e., 2 min and 37 s

over the recent years. Attempts by policymakers to identify, mitigate, and counteract misinformation, disinformation, and propaganda is also not new.

As the COVID-19 pandemic spread across the globe, there was a surge in COVID-related misinformation and scams being spread on multiple social media platforms. These include misinformation regarding face masks, social distancing, lockdowns, and vaccination (e.g., Bill Gates behind vaccines containing microchips to track/control people). Another prominent example is that of misinformation related to potential COVID-19 vaccines in Norway. Figure 3.1 illustrates examples of this misinformation campaign that involved videos that were published on the Bitchute[3] platform which also shows evidence of having links to various social media platforms such as blogs, websites, and Twitter.

The problem of misinformation is actually worse than the pandemic itself, which is why it is often referred to as an “infodemic” or more specifically, a “misinfodemic”. Like the pandemic, disinformation cases are also rising exponentially. These cases are more difficult to track than the epidemic, as they can originate in the dark corners of

[3] BitChute is a video hosting service claiming to put creators first and provide them with a service that they can use to flourish and express their ideas freely. https://www.bitchute.com/.

the internet. To make matters worse, we cannot enforce a lockdown on the Internet to stop the spread of this infodemic, because during crises the Internet is usually the only mode of communication for many information consumers. There have been some quarantine efforts, however. For instance, social media companies such as Facebook and YouTube, and retail companies such as Amazon are doing their best to block such content, suspend bad actors or scammers who are spreading misinformation or trying to profit off of this adversity. But such cases are simply too many and growing too fast. What makes this problem worse is the fact that the information spreads like a wildfire on the internet, especially the false or misinformation. COVID-19 misinformation has caused public fear and anxiety, evidenced by the use of hashtags like #PLANDEMIC (Kearney et al., 2020), #FaceMasksKill, #lockdownskill, #COVIDVaccineKills, and #NoVaccineForMe. Many studies have concluded that misinformation travels faster than its corrective information, and the more questionable the misinformation is the faster it travels. This is simply because on social media people usually have a lot more virtual friends than they do in their real lives. So, if they share or retweet some misinformation, wittingly or unwittingly, they expose all their virtual friends to the misinformation.

There are similarities between misinformation about COVID-19 and other misinformation cases that we have studied for NATO, U.S., EU, Singapore, and Canada. Like in other cases, the motivation for spreading COVID-19 misinformation is monetization or intended to provoke division, confusion, and hysteria. For instance, Russian state media is pushing the conspiracy theory that the virus is manufactured by the western rich and elites to suppress the poor; and Chinese state media is pushing the conspiracy theory that the US Army created the virus to bring the Chinese economy to a grinding halt (Seitz, 2021). And then there are scammers who are selling fake masks, fake cures, and using fake websites to ask for private/sensitive information from people by posing as government websites. However, there is a significant difference between COVID-19 and other misinformation campaigns that we have studied before. Being a global and rapidly evolving crisis, the nature of COVID-19 misinformation is extremely diverse and super-fast. Other misinformation campaigns were specific to an entity, event, region, elections, and military exercises. However, misinformation about COVID-19 has both global as well as regional narratives. While fake masks, fake cures, etc. affect a global audience, the regional narratives include promoting medicines for bovine coronavirus as a cure for the human coronavirus affecting rural/agriculturalist regions. Moreover, the misinformation about COVID-19 ranges from health to policy to religion to geopolitical affairs, i.e., highly topically diverse. Given the volume, velocity, veracity, and variety of COVID-19 related misinformation, research is warranted to study such campaigns and their organization. As resources are stretched thin, government and other regulatory bodies cannot afford to investigate all the misinformation campaigns and scams. Research could help prioritize investigation of misinformation campaigns and scams. We have been gathering data on several COVID-19 misinformation campaigns to forecast themes (Marcoux et al., 2020), contagion of toxic behaviors (Obadimu et al., 2021), epidemiological modeling, and coordination/mobilization (Alassad et al., 2020; Spann et al., 2020). Working with the Arkansas Office of the Attorney

General, we developed a website (cosmos.ualr.edu/covid-19) that tracks conspiracies, scams and false/misleading information. Using cyber-forensics to advance this research, we identified over 600 campaigns, wherein the findings are absorbed by policymakers via reporting to the Arkansas Office of the Attorney General to aid in proactive decision making, while also serving to educate the public and society via our website.

While previous studies have identified strategies including algorithmic manipulation (Sanovich et al., 2018), bots (Agarwal et al., 2017) and trolls (Ananyev & Sobolev, 2017) to influence and polarize crowds, these studies analyze social media platforms in silos and ignore the roles and relations among actors and groups that span platforms. Disinformation campaigns often begin on one platform and then spill over to others and are sometimes executed simultaneously on multiple platforms. Researchers (Agarwal & Bandeli, 2020; Wilson & Starbird, 2020) have identified this problem and have highlighted the need for systematic multiplatform analysis. Aggregating results from individual platforms or combining different platforms using multi-modal network analysis does not provide a holistic view. The supra-dyadic relations among the actors, groups and platforms cannot be explained by conventional social network analysis (SNA). A multiplatform, multi-tiered framework is needed to overcome the shortcomings of conventional dyadic relations analysis. Multiplatform campaign framework development requires data acquisition from various online social media platforms. Complex disinformation campaigns are often run on several platforms, involving hundreds of thousands of users, and millions of information objects (tweets, videos, website links, etc.) in different languages. The data is characterized by the 4Vs of big data (i.e., volume, velocity, variety, and veracity) that requires sophisticated data collection, cleaning, and indexing processes.

3.2 Literature Review

In this section, we discuss the relevant literature regarding misinformation and disinformation from various perspectives. We present this review to highlight the trajectory that misinformation and disinformation has taken historically and up to the current climate within the COVID-19 pandemic. First, we discuss the history of misinformation. Then we delve into the highlights of Russian disinformation. Next, we discuss the use of misinformation within political environments. We then highlight disinformation within the Indo-Pacific region. We then discuss the role media plays in the spreading of misinformation. We then examine the effects of misinformation during COVID-19. Finally, we discuss the various computational approaches to detect disinformation and the need for additional approaches.

3.2.1 History of Misinformation

The use of misinformation has a long history, going back before the pre-printing era. Back in 63 BC, Roman Emperor Augustus Octavian is said to have used disinformation to enable him to be victorious over Marc Anthony (Commonsense.org, 2017). In the sixth-century AD, historians used stone tablets or papyrus to spread misinformation about members of the ruling elite (Burkhardt, 2017). During the post-printing era, however, the spreading of misinformation increased drastically and widely, often aided by the exploitation of the illiterate citizens by those from the literate groups. Satire, rumor, and hoax newspaper articles became common, such as that by Edgar Allen Poe in 1844 claiming that a hot air balloonist crossed the Atlantic in just three days; and those by French writers, such as one reporting that a dragon-like creature was captured in Chile and was being shipped to France. Misinformation was also suspected of streaming into the political landscape of the early eighteenth century as evidenced by an essay written by Jonathan Swift entitled, “The Art of Political Lying.” And it’s quite rare for one to not know a little about the public panic in the United States in 1938 that followed the fake news broadcast by Orson Welles’ of his science fiction book, *The War of the Worlds*. The availability of the internet in the late twentieth century, and later the availability of mobile connected devices, however, provided the means for the spread of misinformation to explode exponentially across the globe. Many journalists and researchers attempt to provide a history of the use of fake news or misinformation on social media, such as the BBC, who asserts that most of the content is political in nature (Wendling, 2018); while others present lists of examples over the years, such as the Social Historian (Starmans, 2018), who discusses examples such as that of the mid-1700s when fake news was printed in Great Britain about King George II in an attempt to weaken his perceived power in the eyes of the people.

3.2.2 Russian Disinformation

Russia’s use of disinformation and computational propaganda has been the focus of numerous researchers over the years. Russia has targeted traditional media (TV, Radio, Newspaper, etc.) and digital disinformation campaigns against several countries including Armenia, France, Germany, and the United States (Sanovich, 2017). RT and Sputnik are two of their most prominent outlets, where they make use of English-speaking anchors in an attempt to increase their traction across the globe. Much of the Russian disinformation targets political topics in order to create division, or exploit already-existing divisions, within the public discourse and societies. The most well-known Russian campaign to date was their involvement in creating divisions before and during the 2016 United States presidential campaign, wherein they made use of a large internet troll farm (a group of paid, internet savvy, or easily trained, youths) to create and spread fake content across multiple social media platforms, often with the aid of in-house created bots (non-human social media accounts) that are programmed to quickly and repeatedly post content on various online social networks (Miller, 2019). Badawy et al. (2018) analyzed this political manipulation on Twitter using a large dataset with over 43 million election-related posts by about 5.7 million distinct users. Anderson and Moody (2020) also studied the Russian influence on the 2016 US presidential election, specifically showing how memes (memetic content) were widely used as a means to create stigma and filter bubbles wherein users tend to "see only what they want to see" and tend to be highly opposed to views from "the other side" or opposing viewpoints on an issue. Russian disinformation is so rampant that commissions and working groups have been organized in several countries in order to study, mitigate, and attempt to counteract it. One such group is the Commission on Security and Cooperation in Europe, which studies the effects of Russian disinformation on countries such as the Ukraine, and other Western alliances such as NATO and the EU (Commission on Security and Cooperation in Europe, 2017). Another such group is the Center for Strategic & International Studies, which studies methods for countering Russian disinformation in countries such as The Czech Republic and Estonia, as well as more generally against NATO and the EU (Center for Strategic & International Studies, 2020). Russian disinformation campaigns tend to unfold in a multi-stage manner, which involves first infiltrating an information space or audience, then attempting to influence it, and then introducing damaging narratives against specific targets such as political figures or organizations (Watts, 2018). One working group based out of Estonia, the International Centre for Defence and Security (ICDS), has been persistently assessing Russia's disinformation campaigns relating to the COVID-19 pandemic. The ICDS has discovered that during the pandemic, Russia's tactics have diverged somewhat from their more traditional techniques to include a greater emphasis on the manipulation of debates on social media (Weitz, 2020). Evidence of Russian disinformation campaigns is so robust that even mainstream media outlets report on it, such as NPR (Allyn, 2020), who discussed the evidence of Russian operatives expanding across 300 social media platforms over the past six years.

3.2.3 Misinformation in the Political Environment

As can be expected from the previous discussions so far, much research has been focused on the use of misinformation within the political realm. Mason et al. (2018) discussed the importance of media literacy education within democracies so that citizens can be properly informed about strategies for deciphering misinformation from facts. The authors emphasize that this education must include K-12 but also expand into adult contexts. Valenzuela and colleagues (2019) pointed out that the increasing use of social media by citizens in order to consume news is a type of paradox in that it increases their political engagement while also increasing their exposure to misinformation as well as increases their potential to share misinformation. Jerit and Zhao (2020) discussed how misinformation infiltrates political systems on a grand scale and is increasingly difficult to correct, especially within the United States. The authors also touch on the concept of conspiratorial beliefs, and how they are often rooted in psychological predispositions that make them extremely difficult, if not oftentimes impossible, to correct or change. In general, however, political misinformation, once it takes hold in the information consumer's mind, is continually led by their motivation to strive to some end goal, such as being self-confident in their knowledge about an issue. This motivation can be so strong that it often prevents them from actively seeking out and evaluating evidence that might change their beliefs. The McCourt School of Public Policy (2020) continually studies the impact of misinformation spread via social media on elections. They also paid special attention to the impact social media misinformation had on the 2016 U.S. presidential election and asserted that social media platform policies were severely lacking at the time in terms of identifying, mitigating, and preventing the spread of blatant misinformation.

3.2.4 Misinformation in Indo-Pacific Region

The study of the effects of misinformation on Indo-Pacific countries is also growing. Fernandez of Global Risk Insights discussed how rumors and fake news in Asian politics has been a problem over the past decade, and that government initiatives have so far failed in its curtailment (Fernandez, 2019). Palatino, with advocacy group IFEX (Palatino, 2019), has also studied the issue of misinformation in the Indo-Pacific region. Palatino points out the drastic effects of misinformation about the Rohingya minority in Myanmar, and how oftentimes governments in Indo-Pacific countries resort to imposing communication and internet blackouts in an attempt to reduce the visibility of the events. Such blackout tactics were reportedly used in Myanmar, Sri Lanka, Taiwan, Kashmir, Thailand, the Philippines, and others. Cha and colleagues (2020) discussed the importance of detecting fake news in social media specifically within the Indo-Pacific region. The authors highlighted examples such as the spreading of false rumors regarding the 1986 nuclear accident in Japan wherein edited images were circulating on social media claiming to show radioactive material spilling into the Pacific Ocean. The authors argue that although many countries such as South Korea, Singapore, Japan, Taiwan, the Philippines, Cambodia, and Malaysia have enacted laws aimed at preventing the spread of misinformation on social media, the problem remains rampant. As the COVID-19 pandemic spread into the Indo-Pacific region, misinformation about food safety began to spread (Clark, 2020). The Food and Agriculture Organization of the United Nations (FAO), among other organizations, have attempted to quell false claims of COVID-19 tainted food that have spread across Asia and the Pacific (Food & Agricultural Organization of the United Nations, 2020).

3.2.5 Role of Media in Misinformation

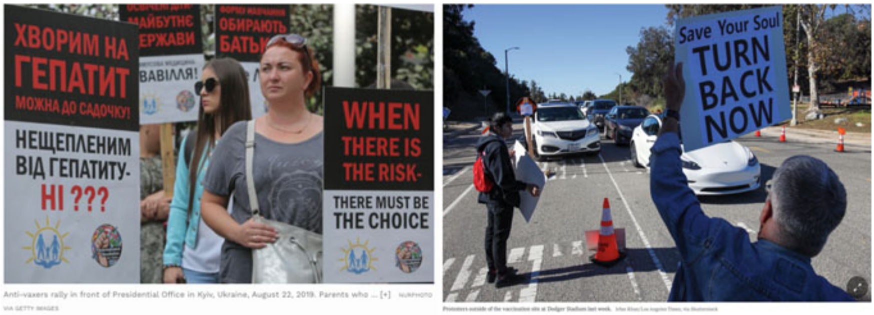
The role that mainstream media plays in the spread of misinformation has been studied by several researchers over the years. Mihailidis and Viotty (2017) argue that mainstream media outlets have been seemingly willing to actively participate in the spread of misinformation, and the effects have been to polarize information consumers, perpetuating already existing partisan divides within societies. They argue that media literacy campaigns are key to increasing the critical thinking capabilities of media consumers. Tsfati and colleagues (2020) also assert that mainstream media plays a direct role in the dissemination of disinformation and fake news. They contend that mainstream media outlets have the ability to reach larger proportions of information consumers and, therefore, more attention is dedicated to these falsities, even when the media coverage is an attempt to correct the misinformation. The authors give an example of the "Pizzagate" term, which refers to the spread of a conspiracy theory that falsely claimed that a pizza restaurant in Washington, D.C. was the meeting place for several high-profile business leaders and political figures falsely believed to be engaged in child sex trafficking. Coverage of the conspiracy theory was frequent on several mainstream media outlets such as the BBC, NBC, The New York Times, Rolling Stone, among numerous others. The authors state that even when mainstream media outlets report on fake news items in order to attempt to correct them, the consumer is more likely to remember the disinformation, misinformation, rumor, etc., because it increases the familiarity of the wrong information and therefore the consumer is more likely to retain it.

Galeano and colleagues (2019) showed how mainstream media played a large role in the influence of the 2019 Canadian Federal election. The researchers analyzed a large dataset of election related YouTube videos and user comments on those videos that allowed them to identify a set of key players who increased their targeted posting frequency around prominent events such as the SNC-Lavalin investigation, which was a political scandal involving incumbent candidate Justin Trudeau. Among the key actors attempting to influence the minds of voters were "Rebel News", a Canadian political and social commentary media outlet. "Global News" was another dominant actor, as well as "2 Average Dudes" and "CBC News". "2 Average Dudes" is a popular YouTube channel that continually communicates anti-Trudeau messaging. The channel includes several "bridge nodes" (users in the social network) that connect other users to another conservative, anti-Trudeau Canadian YouTube channel called "True North". Social cyberforensics conducted by the researchers revealed that Rebel News was engaged in widespread influence and information maneuvering tactics leading up to the election, including the use of a network of at least 289 websites whose links were shared across the YouTube network.

3.2.6 Misinformation During COVID-19

An abundance of researchers have put forth efforts to study the spread of misinformation during the global COVID-19 pandemic that began in 2019. Banerjee and Rao (2020) described it as a “digital infodemic” wherein pervasive misinformation items related to various aspects of the virus have caused mass anxiety, fear, and uncertainty worldwide. This misinformation includes fake treatments and cures, prejudice against certain ethnicities and nationalities, stigma against people who wear face masks, and campaigns to try to influence citizens to resist complying with precautionary measures such as social distancing and staying home. Bao et al. (2020) addressed the role of digital media during the pandemic. The authors emphasize the importance of improving trust in digital media by reducing the spread of misinformation and educating information consumers on how to discern between real and fake news, while also continuing to protect user privacy. The authors point out, however, that there is still a lack of formalized guidance as to how social media platforms and other digital media tools such as Zoom (a leading video communications platform) should do this. Chou and colleagues (2020) proposed a framework for how health misinformation on social media can be better identified and how policies and practices can be improved. They first suggest the need for enhanced surveillance, especially for thus far understudied platforms such as WeChat, Tumblr, Reddit, and Pinterest. This includes the need to gain a better understanding of how images, memes, and videos are being used across social media platforms. The authors also suggest a need for the examination of the role that emotion, cognition, and identity play in relation to the spread of health misinformation. For example, what are the psycho-social-emotional drivers that lend to the acceptance of misinformation among information consumers? Of further importance, is the need for researchers to communicate their findings to policymakers, not only those within social media platforms, but also within public health and governmental agencies. Shahi et al. (2020) conducted an exploratory study of COVID-19 misinformation on Twitter. The researchers concluded that tweets containing misinformation during their time frame of analysis (January to mid-July of 2020) were more often communicating some type of discreditation attempt against some other information, and that false claims disseminated faster across the platform than did partially false claims.

Misinformation about COVID-19 vaccines is rampant across social media, and policymakers must do more to detect, mitigate, and prevent such harmful content (Horton, 2020). Anti-vaccine rhetoric online is deterring many citizens across the globe from obtaining the COVID-19 vaccine (Loomba et al., 2020). As mentioned in Sect. 3.1, anti-COVID-19 vaccine misinformation includes false reports of deaths occurring after receiving the vaccine, as well as various conspiracy theories related to the vaccine (Reiss, 2021). This COVID-19 anti-vaccine misinformation has a global impact (Fig. 3.2) (Fernandez, 2021; Salzberg, 2020), and it has even led some pharmacists to deliberately sabotage COVID-19 vaccines (Lee, 2021). Although online social networks such as Facebook have claimed to have actively attempted to remove such misinformation, the problem still persists (Bond, 2021).


ХВОРИМ НА ГЕПАТИТ МОЖНА ДО САДОЧКУ!
НЕЩЕПЛЕНИМ ВІД ГЕПАТИТУ- НІ ???
ОБИРАЮТЬ БАТЬК
WHEN THERE IS THE RISK- THERE MUST BE THE CHOICE
Save Your Soul TURN BACK NOW

Anti-vaxers rally in front of Presidential Office in Kyiv, Ukraine, August 22, 2019. Parents who ... [+] NURPHOTO VIA GETTY IMAGES

Protesters outside of the vaccination site at Dodger Stadium last week. Irfan Khan/Los Angeles Times, via Shutterstock

Fig. 3.2 Example vaccine hesitancy demonstrations

3.2.7 Computational Approaches to Detect Misinformation

The popularity of social media platforms has increased exponentially in recent years encouraging many academics, marketing agencies, and cybersecurity firms to research and explore the information, behaviors, and impacts generated by these online platforms (Myers et al., 2014). Online social networks have experienced a surge in negative behavior campaigns such as fake news, misinformation, and disinformation generated by malicious users to achieve their desires. Spreading this type of real-time misleading information to millions of online users has resulted in many negative consequences to politics, economic issues, social behaviors, and people's daily life (Zhang & Ghorbani, 2019).

Research has been conducted in network structures, intelligence, and tools to outline, distinguish and compartmentalize fake news, misinformation, and disinformation campaigns (Søe, 2018). However, the authors concluded that current algorithms are insufficient and need additional capabilities to handle such topics. There are many fact-checking resources available in terms of news related to politics, technology, business, civic life, social media, and emails (Zhang & Ghorbani, 2019). For example, Hoaxy.iuni.iu.edu (Shao et al., 2016) is a framework to check misinformation related to various sources of information on social media, websites, and emails. This website relies on other known reputable websites to check the content, tweets, and any claims about those topics (Zhang & Ghorbani, 2019). Resources such as Factmata.com, Hoax-slayer.com, PolitiFact.com, and Snope.com are designed to fight online fake news topics, but fighting online fake news, misinformation and disinformation is still a difficult task (Shao et al., 2016; Søe, 2018).

Many efforts have been made to identify and prevent users that participate in misinformation or disinformation campaigns and spread fake news, where suspending these users early, would minimize the negative effects generated by such malicious users. Shu et al. (2017) suggested ways to minimize online fake news by suspending or removing suspicious online accounts, but also stressed the need to maximize online true news. Cha et al. (2020) discussed various recent deep learning techniques that attempt to further improve performance and interpretability of fact-checking. These methods include: Feature-engineering, Matching-based, Representation-learning, and Multimodal. Each of these models utilize large text datasets to build classifiers for labeling false content with the intent to automatically detect the false content when the models are applied to new text datasets. Continued research and fine-tuning of such models are still needed, however, and collaboration among researchers and policymakers is of great importance. In the next sections we will highlight our approaches to overcome some of these challenges.

3.3 Data Collection Method

Our data collection methodologies can be separated into two main phases: (1) data acquisition and (2) data analysis and model development. The main tasks of Phase 1 (Fig. 3.3) include identifying keywords, events, cyber incidents, selecting reliable and optimal social media sources, and collecting the information from relevant social media sources and metadata using cyber forensics. The data is largely unstructured and noisy which warrants cleaning, standardization, normalization, and curation before proceeding to Phase 2.

Phase 2 entails categorizing misinformation cyber incidents, correlating the identified incidents to current news articles, and examining the social and communication networks to identify prominent actors and groups. Research in this phase also often reveals the coordination strategies and tactics, techniques, and procedures (TTPs) used by these actors and groups. Additionally, a multilayered network analysis approach (Fig. 3.4) is considered to model multi-source, supra-dyadic relations, and shared affiliations among deviant groups. Deviant groups are known to coordinate their acts using one or more social media platform (Agarwal & Bandeli, 2020).

Our experience with previous misinformation studies and guidance from policymakers has helped us develop newer methodologies to handle challenges with

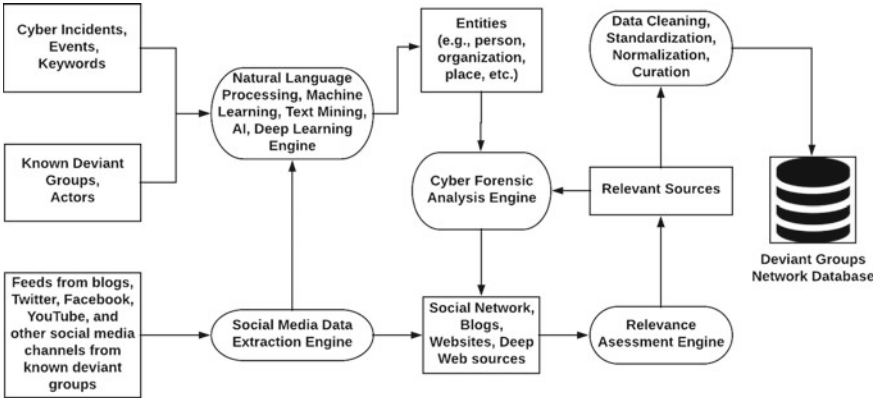

Cyber Incidents, Events, Keywords
Known Deviant Groups, Actors
Feeds from blogs, Twitter, Facebook, YouTube, and other social media channels from known deviant groups
Natural Language Processing, Machine Learning, Text Mining, AI, Deep Learning Engine
Social Media Data Extraction Engine
Entities (e.g., person, organization, place, etc.)
Cyber Forensic Analysis Engine
Social Network, Blogs, Websites, Deep Web sources
Data Cleaning, Standardization, Normalization, Curation
Relevant Sources
Relevance Asessment Engine
Deviant Groups Network Database

Fig. 3.3 Social media data collection and curation methodology

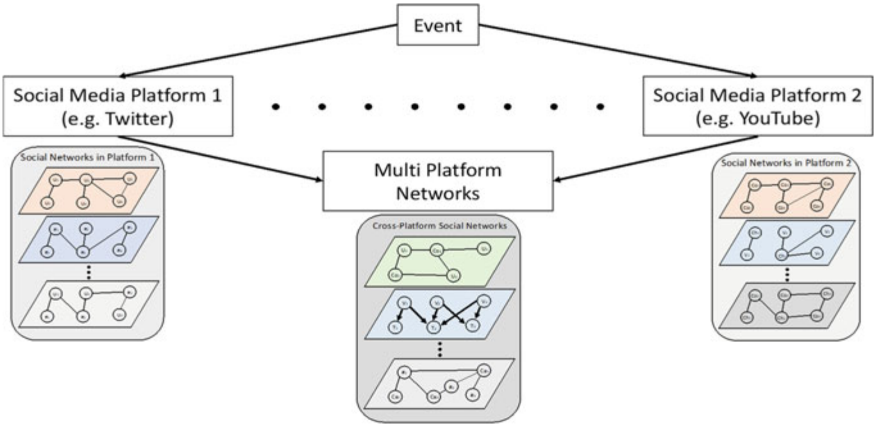


Fig. 3.4 Multilayer network analysis of deviant cyber flash mobs' social media communications

collecting misinformation cases for COVID-19. We are gathering data using a three-pronged approach: (1) finding instances of myths, rumors, disinformation, and misinformation surrounding the keywords "coronavirus" and "COVID-19" that are being spread on various social media platforms such as Twitter, Facebook, YouTube, WhatsApp, Telegram, and basic text messaging services; (2) finding credible sources that have information that debunks the false claims or false narratives; and (3) finding, when possible, the social media sources that are spreading the false information. The data capture methods that have proven to be most useful in our research include data crawling using APIs. Additionally, we have developed several in-house data collection and processing tools for larger-scale datasets. These include YouTubeTracker (https://cosmos.ualr.edu/tools/youtubetracker/) and Blogtrackers (https://cosmos.ualr.edu/tools/blogtrackers/), which are freely available for policymakers, regulators, and the general public. We are continuously identifying new cases of fake websites, misinformation content, and bad actors. We are using our social network analysis and social cyberforensic methodologies to identify such cases. Furthermore, this is not a one-time process. We repeat it every day to ensure that we can capture as many new cases as possible. The next sections will give a brief high-level technical overview of these data collection methodologies.

3.3.1 Twitter Data Collection

To capture COVID-19 misinformation data from Twitter, we use the Twitter API to collect co-occurring hashtag networks. We can search for an online campaign by using hashtags, or a text query. For example, our study on protestor orchestration (Spann et al., 2020) examines connective action in the communities using #MichiganProtest, #MiLeg, #Endthelockdown, and #LetMiPeopleGo hashtags over a 50-day period from April 01, 2020 to May 20, 2020. During this timeframe, state and local stay-at-home orders were issued due to the novel coronavirus Pandemic. The data collected resulted in a network of 16,383 tweets, with 9985 unique users. Using this data, we created a communication network between users consisting of the retweets and mentions. The graph revealed 3632 nodes with 382 communities. The text data from the tweets was also used to train a Latent-Dirichlet Allocation (LDA) model to generate semantic themes in each community, which allowed us to highlight both shared social interests among communities and polarizing themes across each community as they emerged.

3.3.2 YouTube Data Collection

To understand the online toxicity surrounding the COVID-19 pandemic, we employed a methodology that consists of four components: (1) data crawling and processing, (2) toxicity analysis, (3) topic modeling, and (4) social network analysis. We leveraged the YouTube search Data API to extract channels, videos, and comments using the following keywords: "coronavirus", "corona", "virus", "COVID19", "COVID", and "outbreak". Our dataset consisted of 544 channels, 3488 videos, 453,111 commenters, and 849,689 comments (826,569 of which were unique comments). The time frame of our dataset spans the period from January 2020 through early May 2020. The comments were primarily from videos that were categorized as "News & Politics" (94%), followed by "Entertainment" (0.8%). To reduce noise in the extracted data, several data processing steps were performed including data formatting, data standardization and data normalization. We combine the techniques of topic modeling, toxicity analysis, and social network analysis to emphasize the utility of this multi-methodology approach for policymakers to identify misinformation and reduce hate speech.

3.3.3 Misinformation Data Collection

To properly characterize which stories were misinformation and which were not, we took additional steps to validate the context of the narratives found within our datasets. Initially, the misinformation stories in our datasets were obtained from a publicly available database created by EUvsDisinfo in March of 2020. EUvsDisinfo's database, however, was primarily focused on "pro-Kremlin disinformation efforts on the novel coronavirus". Most of these items represented false narratives that were communicating political, military, and healthcare conspiracy theories in an attempt to sow confusion, distrust, and public discord. Subsequently, additional misinformation stories were continually gleaned from publicly available aggregators, such as POLITIFACT (https://www.politifact.com/), Truth or Fiction (https://www.truthorfiction.com/), FactCheck.org (https://www.factcheck.org/), POLYGRAPH.info (https://www.polygraph.info/), Snopes (https://www.snopes.com/), Full Fact (https://fullfact.org/), AP Fact Check (https://fullfact.org/), Poynter (https://www.poynter.org/), and Hoax-Slayer (https://www.hoax-slayer.com/). The following data points were collected for each misinformation item: title, summary, debunking date, debunking source, misinformation source(s), theme, and dissemination platform(s). The time period of our data set is from January 22, 2020 to February 11, 2021. The data set consists of 612 unique misinformation items. For many of the items, multiple platforms were used to spread the misinformation. For example, oftentimes a misinformation item will be posted on Facebook, Twitter, YouTube, and as an article on a website. For our data set, the top-used platforms used for spreading misinformation were websites, Facebook, Twitter, YouTube, and Instagram, respectively. Next, we discuss the research methodologies used in this research.

3.4 Research Methodologies

We use numerous data analysis frameworks in our research on misinformation to help inform our decisions. These include topic modeling, social network analysis, social cyberforensics, toxicity analysis, and epidemiological modeling. Each of these frameworks are discussed below.

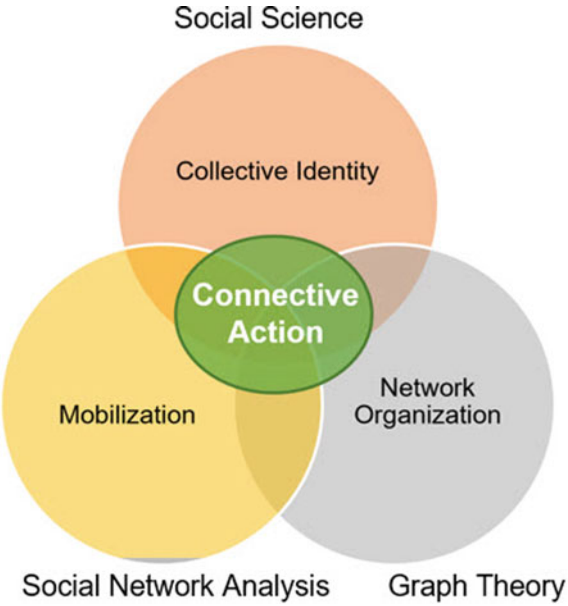
3.4.1 Topic Modeling

We have collected a growing corpus of COVID-19 misinformation narratives from various social media (YouTube, blogs, Twitter, etc.) and other outlets (text messages, emails, WhatsApp, etc.) and developed a tool for visualizing the evolution of specific topic themes over time (Marcoux et al., 2020). We use a two-step methodology to identify and deduce these topics. First, through a manual curation process, we aggregate different misinformation narratives for later processing. We consider misinformation narratives as any narrative pushed through a variety of outlets (social media, radio, physical mail, etc.) that has been or is later believably disproved by a third party. This corpus constitutes our input data. Secondly, we use this corpus to train an LDA topic model to derive lexical meaning from the text and generate subsequent misinformation themes for analysis.

3.4.2 Social Network Analysis

Misinformation stories that get propagated on social media often lead to protests, where actors come together under collective action to pursue change. The protests can also result in deviant cyber flash mobs (DCFMs) which can produce cybernetic events, where the protests become physical. Social network analysis provides several techniques to measure key metrics and provide insights to policymakers to help understand these connective action movements. Graphing the network provides a map of the users participating in the campaign. This social network map can help researchers and policymakers determine the structure and flow of information between individuals within groups, such as who was sustaining the movement or directing messages to other users. When we consider digitally enabled protests, they need to have three key features to form a successful connective action: they should have collective identity, they should produce organization, and they should mobilize participants (either online or physically) as shown in Fig. 3.5. The type of organization, whether organizationally enabled or self-organizing, will largely determine whether the movement is a collective action or connective action network. Connective action movements are often self-organized and include self-expression action frames (Segerberg & Bennett, 2013). The organization and action frames can be measured through our connective action framework. Social network analysis, Natural Language Processing (NLP), DCFM modeling, and social cyberforensic

Fig. 3.5 Elements of connective and collective action



techniques (Al-khateeb & Agarwal, 2019) were used to develop and operationalize the theoretical frameworks used in our studies. We use these techniques to identify user and community roles in the network as they relate to the types of affordances offered by social media and digital technologies.

3.4.3 Social Cyberforensics

We define social cyberforensics (SCF) as a branch of cyberforensics. SCF is the process of investigating the relationships among “entities” on social media and revealing the digital connections among them by extracting and collecting metadata associated with their social media accounts (Al-khateeb & Agarwal, 2019). We use SCF to study cross-platform affiliations and uncover hidden relationships among various online groups or actors. Mix-media and cross-media dissemination are common tactics used to spread misinformation and disinformation. A mix-media information dissemination campaign refers to the use of multiple social media channels to diffuse a narrative. More precisely, the information campaign can be observed on multiple social media sites with text, images, audio and video content. For example, blogs are used for framing the narratives on other media sites, while Twitter may be used for dissemination and discourse development. The content may not be identical on the various social media channels, but it pertains to a particular information campaign. SCF methods help us to uncover these dissemination tactics by identifying the affiliations of the user, geolocation, IP address, and other digital presence. These methodologies have been thoroughly tested during previous cyber propaganda disinformation campaigns that were projected against NATO forces, such as Trident Juncture (Galeano et al., 2020).

3.4.4 Toxicity Analysis

High levels of toxic content prevent the dissemination of important and time-sensitive information and jeopardizes the sense of community that online social networks seek to cultivate. Another step in our methodology to detect misinformation is to compute toxicity scores for comments on YouTube datasets. In our toxicity research, Obadimu et al. (2021) developed techniques to analyze toxic content and actors that propagated the content on YouTube during the initial months after COVID-19 information was made public. The toxicity model uses artificial intelligence and machine learning to determine whether a comment could be perceived as “toxic” to a discussion. Since the health of the social media user is at risk of being negatively affected by the toxicity that is being interjected into social media platforms, and due to potential power of the YouTube platform to be used as a vehicle for the propagation of toxicity, it is imperative that policymakers and regulators are able to identify this toxicity and the commenters that attempt to propagate it. Our research utilizing COVID-19 as a case study is able to identify and profile toxicity within the network, examine themes among toxic comments, investigate connections among toxic commenters, and simulate the removal of these commenters to discover the impact on the health of the network.

3.4.5 Epidemiological Modeling

Characterizing information diffusion on social platforms like Twitter enables us to understand the properties of underlying media and model communication patterns. Similar to how an infectious disease spreads in a community, misinformation can spread through social networks. Our emerging research on epidemiology in the information environment enables us to model misinformation contagion processes in social networks. Epidemiological models divide the population into different compartments that represent the state of each user involved in the social network. The SIR model, also known as the Kermack-McKendrick model (Kermack & McKendrick, 1927), is one of the most frequently used epidemic models. This model typically tracks the flows of people between three compartments: susceptible (S), infected (I), and recovered (R). The SEIR model is used for information diffusion introducing an Exposed (E) state where individuals have a period of incubation before getting infected by others (Bettencourt et al., 2006). The SEIZ (susceptible, exposed, infected, skeptic) model was also used for modeling the spread of news and rumors on twitter (Pavlopoulos et al., 2019). One major limitation of the SIS (susceptible-infectious-susceptible) and SIR (susceptible-infectious-recovered) epidemic models is that when a susceptible person is in contact with an infected one, there is just one possibility, which is that the user can only move to the infected compartment ("Compartmental models in epidemiology," 2021). This assumption does not apply well to the spread of misinformation, especially on social media, and specifically on Twitter. Individuals can have different complicated beliefs when they are exposed to items of misinformation on social media. Some people may have different viewpoints about the misinformation item; others may need some time to come to believe it; or others can be skeptical as to the accuracy of what they saw. When people are faced with misinformation on social media, they may be convinced to spread the misinformation after thorough consideration, which requires some time for some people, while being immediate for others. In addition, there is a possibility that some individuals that are exposed to an item of misinformation never show any reaction to it. These possibilities are not covered by the SIS and SIR models. Based on the provided reasons, we decided to use the SEIZ model, which is a more powerful model and more applicable to the spread of misinformation on Twitter. In the next section, we present the various findings from our numerous studies.

3.5 Research Findings

In order to show how the administrators of online social networks can potentially improve the overall health of their networks, we continue to examine misinformation themes from across multiple social media platforms using a broad spectrum of quantitative and qualitative research methods. The following sections present some practical findings from our recent studies.

3.5.1 Misinformation Topic Themes

With our research, we have been able to assist policymakers by identifying and quantifying topic trends in misinformation narratives. Our analyses show indications of periodicity in these trends, which can lend to predictive capabilities over time (see Fig. 3.6).

Our research into topic analysis has highlighted some of the narratives that surfaced during the COVID-19 pandemic (Marcoux et al., 2020). From January 2020 to July 2020, we collected 243 unique misinformation narratives and proposed a tool to observe their evolution. We have shown the potential of using topic modeling visualization to get a bird's eye view of the fluctuating narratives and an ability to quickly gain a better understanding of the evolution of individual stories. We have seen that the tool is efficient to chronologically represent actual narratives pushed to various outlets, as confirmed by the ground truth observed by our misinformation curating team and independent international organizations. Working with the

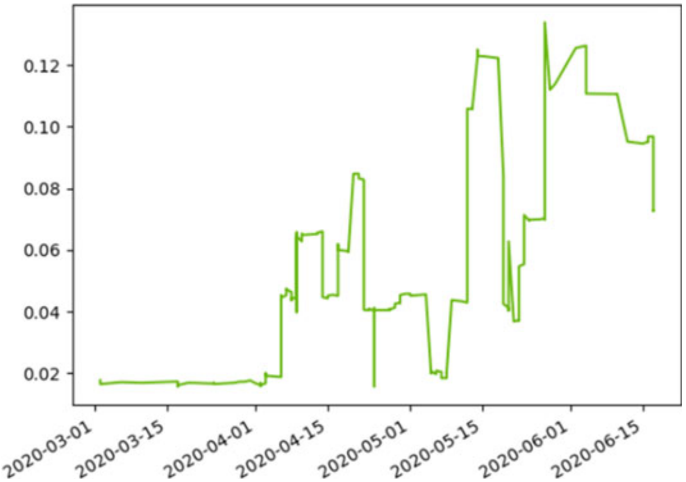


Fig. 3.6 Evolution of a dominant recurring COVID-19 misinformation narrative topic that includes keywords: "hydroxychloroquine", "health", "scam"

Arkansas Office of the Attorney General, this study illustrates a relatively quick technique for allowing policymakers to monitor and assess the diffusion of misinformation on online social networks in real-time, which will enable them to take a proactive approach in crafting important theme-based communication campaigns to their respective citizen constituents. We have made most of our findings available online to support this effort. We then scaled up our data and repeated our methodology on social media data: YouTube video titles and comments. Over concerns of our LDA topic model becoming difficult to scale, we also experimented with an HDP (Hierarchical Dirichlet Process) model, which attempts to infer the number of topics. Although this technique produced slightly less precise results, we noticed that HDP was able to isolate a probable subset of polarizing comments. One possible way around the limitations of this approach would be to use HDP to identify these subsets, filter out irrelevant comments, then apply the LDA model. This may reveal various narratives, some of them spreading misinformation, and further automate the process of identifying online misinformation in uncontrolled spaces.

3.5.2 Misinformation Spread—Epidemiological Modeling

We have applied (Maleki et al., 2020) an epidemiological model to explain how a specific piece of politically-charged misinformation was able to be quickly propagated across Twitter even though it was entirely false. In a misinformation case study about the Washington D.C. blackouts, we specifically evaluated the dissemination of misinformation about the extent of the unrest in Washington, D.C., the capital of the United States. This misinformation item claimed that there was a communication blackout in Washington D.C. because of riots supposedly occurring on Monday, June 1, 2020. This event was a culmination of the "Black lives Matters" protest that sprouted due to the death of George Floyd. As a result, there was a widespread misinformation campaign regarding the communication outage that supposedly happened on Monday, June 1, 2020 in D.C. This was propagated by the use of the #DCblackout hashtag by various actors across the US on Twitter. Numerous Twitter accounts shared photos of a large out-of-control fire close to the Washington Monument. However, other accounts mentioned that the photos were appropriated from the TV show "Designated Survivor" and were not related to the riot and unrest at all. Our methodology involved initial data collection, followed by the subsequent application of the SEIZ epidemiological model discussed in Sect. 3.4.5. We showed how the spread of misinformation concerning the unrest in D.C. propagated on Twitter can be modeled by applying the SEIZ epidemiological model. Acquiring a relative error metric of 0.019 between the real number of tweets and the fitted model demonstrates that the SEIZ model is accurate in evaluating the spread of misinformation on Twitter (see Fig. 3.7). Using this mathematical model for the spread of an item of misinformation can provide an opportunity to evaluate trends. This evaluation can help provide informed strategies for policymakers trying to control the spread of misinformation on social platforms.

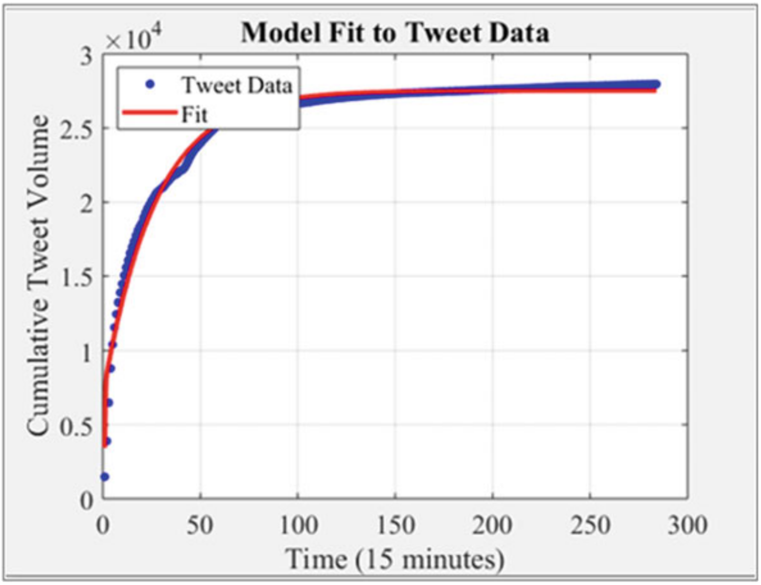


Fig. 3.7 SEIZ Epidemiological model comparing actual number of tweets to predicted tweets in ‘Infected’ compartment for the #DCblackout misinformation hashtag

Applying the SEIZ epidemiological model to social network messaging, we can conclude that all users are initially in the susceptible group. Our results show that the exposed individuals have been transferred to the infected compartment because of self-adoption and not so much because of direct contact with infected users. In other words, surprisingly, other mediums had more influence in a user’s decision to spread an item of misinformation on Twitter than their coming into contact with the tweets about that misinformation.

3.5.3 Connective Action—Protest Orchestration

Our emerging socio-technical research on connective action uses social network analysis, natural language processing, and social science concepts to identify the properties of connective action and how it differs from collective action. We have created a theoretical framework (Agarwal et al., 2012) for operationalizing connective action theory using computational methods grounded in social network analysis. Bridging the gap between social science concepts and computational techniques provides insights to policymakers about online connective action events such as the type of messaging within the campaign, the powerful brokers that are influencing messaging, and whether the campaign has reached critical mass.

Our framework highlights the social science and behavioral theories of affordances, interdependence, and resource mobilization in social networks through critical mass theory. Social movement activists and scholars often use “critical mass” in a loose metaphorical way to refer to an initial group of protesters or actors that is big enough to accomplish social change. As such, our Michigan protest network represents a good case for assessing the use of digital technologies to engage and mobilize citizens. In our case study we analyzed the mobilization effect by examining network modularity and network power over time. We see that as the protest movement gains users, we also see the network begin to separate into homogenous communities (see Fig. 3.8) as they share similar social constructs. The emergence of the network begins to take on an s-shaped function as we look at the collective power of the users in the campaign. In many of the misinformation campaigns, there is an initial growth stage, accelerating stage, and then a decelerating stage. We can also help policymakers by identifying the users associated with these phases and classify them into 3 emergent roles: initiators, amplifiers, and supporters. Initiators may pay the startup costs by assuming the risks associated with being the initiator of a new connective action

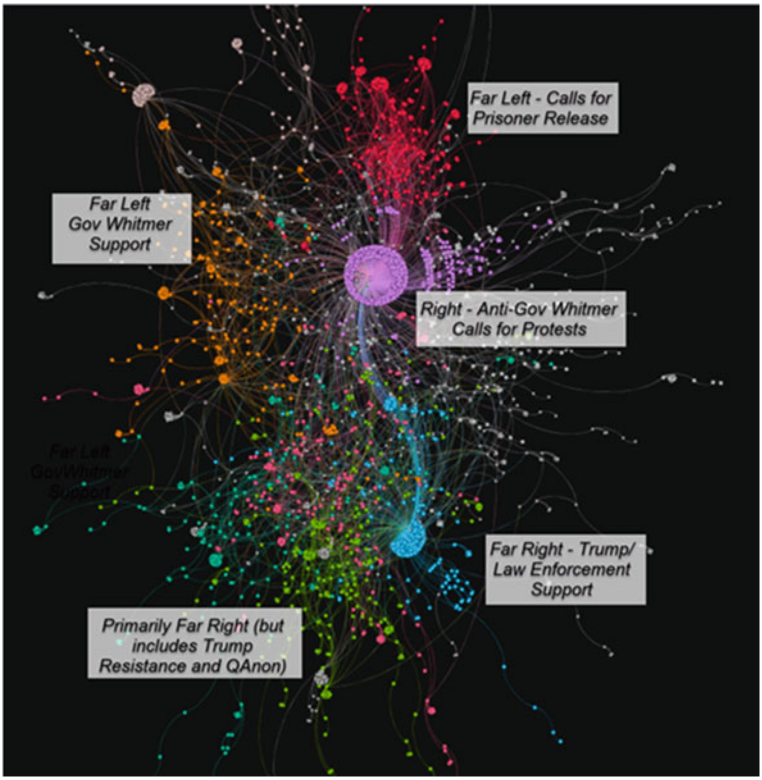

Far Left - Calls for Prisoner Release
Far Left Gov Whitmer Support
Right - Anti-Gov Whitmer Calls for Protests
Far Left GovWhitmer Support
Far Right - Trump/ Law Enforcement Support
Primarily Far Right (but includes Trump Resistance and QAnon)

Fig. 3.8 Top Influential Nodes by high betweenness centrality and outdegree in Michigan anti-lockdown/anti-quarantine protest network

campaign. These users are typically heavy users with high DCFM power and calls to action to encourage others to participate. The initiators create the conditions for widespread contribution of others. Amplifiers accelerate the messaging by providing more inputs to scale the campaign and build momentum. They act by disseminating and circulating content in the information environment, primarily through mentions and retweets, to sustain messaging. Finally, supporters follow the movement and support the connective action activities by retweeting messages of others, with little original content.

Considering the Michigan anti-lockdown protest as a case study (Spann et al., 2020), there are multiple inferences that can be drawn from our analysis. First, our technique suggests that the Michigan anti-quarantine protests were primarily driven by conservative party interests. This is consistent with the physical protest demonstrators that showed up at the Michigan State Capitol supporting their 2nd amendment rights. Figure 3.8 also shows diverse polarization in topics and responses to each category around the protests. The conservative communities appear to have more densely connected communities than the liberal communities. We infer from this data that connective action from these communities is more cohesive and thus more successful in their protest efforts. More broadly, our study demonstrates a way to assess the power of digital communities in mobilizing and coordinating protests. Media reports validate that armed protesters physically protested at the capital resulting in a shutdown of the State Capitol. Finally, our deviant cyber flash mob (DCFM) detection method is a strong indicator of influence power in the network as the campaign emerges. As such, analyzing the follower network of a user does a good job characterizing the concept of interdependence among nodes.

3.5.4 Toxicity and Hate Speech Contagion

The health of the social media user is at risk of being negatively affected by the toxicity that is being interjected into social media platforms. Toxic discourse is known to increase polarization among digital communities. This toxicity contagion has been identified among the commenters on COVID-19 YouTube videos. Due to potential power of the YouTube platform to be used as a vehicle for the propagation of toxicity, it is imperative that we develop techniques to identify this toxicity and the commenters that attempt to propagate it. Based on our studies using large datasets (Obadimu et al., 2021), we have been able to identify a trend in the contagion of toxicity. Through this work, we demonstrate not only how to identify toxic content related to COVID-19 on YouTube and the actors who propagated this toxicity, but also how social media companies and policymakers can use this work. Our research shows how toxicity traverses within a network and affects non-toxic commenters. Commenters with similar toxic levels tend to stay together and replies to toxic comments often were at a similar toxicity level to the original comment. Commenter groups also worked collectively to form echo chambers to amplify toxic beliefs as described below.

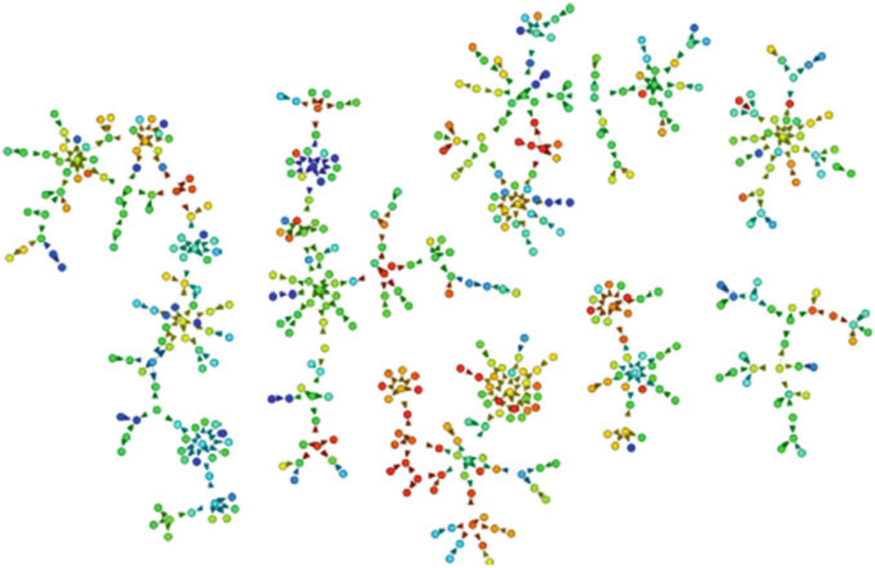


Fig. 3.9 Commenter-reply directed network clusters depicting segregation among commenters based on toxicity.[4]

Figures 3.9 and 3.10 connect to show that the network is not only segregated but there are also groups that exist who are collectively forming echo chambers to amplify toxic beliefs reinforced by communication inside a closed system.

We also observed multiple different subclusters within the same component (Fig. 3.10, circled areas), which indicates, "network homophily" or assortativity based on the toxicity scores. Homophily simply means 'birds of a feather flock together'. In network science, homophily refers to how similar nodes have a higher tendency to group together than do dissimilar ones based on node attributes (toxicity in this case). This work is unique in that we devised a set of computational experiments to show how if social media platforms eliminate certain toxic users, they can improve the state of online discourse by reducing toxicity and hate speech to heal fractured digital communities. Next, we present how we transitioned our findings into a COVID-19 Misinformation Tracker tool and how it assisted the Arkansas state government in raising awareness about the problem of COVID-19 misinformation.

[4] Colors denote toxicity scores from lowest, 0.5 (blue) to highest, 1 (red). Since toxicity scores are based on a probability score of 0 to 1, we focused our analysis on toxic comments that are 0.5 or greater in order to gain a deeper understanding of high toxic content.

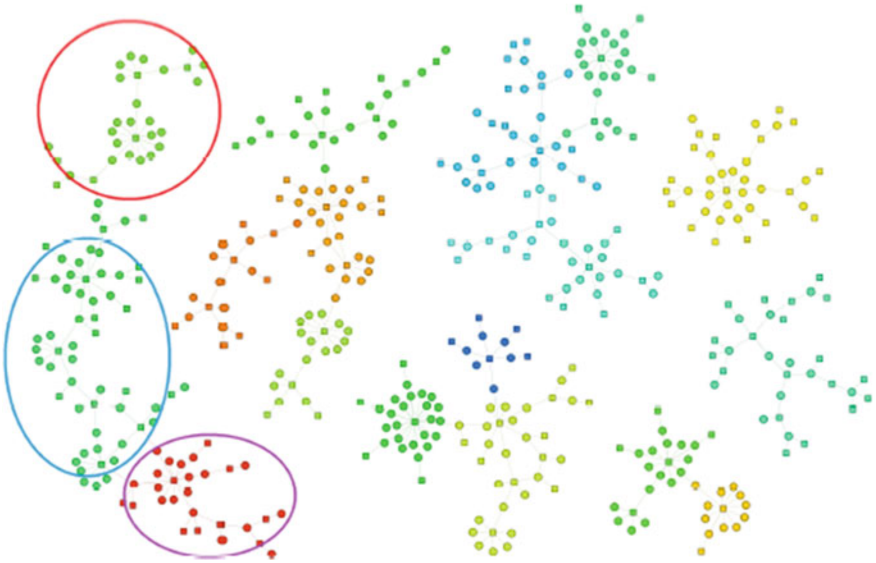
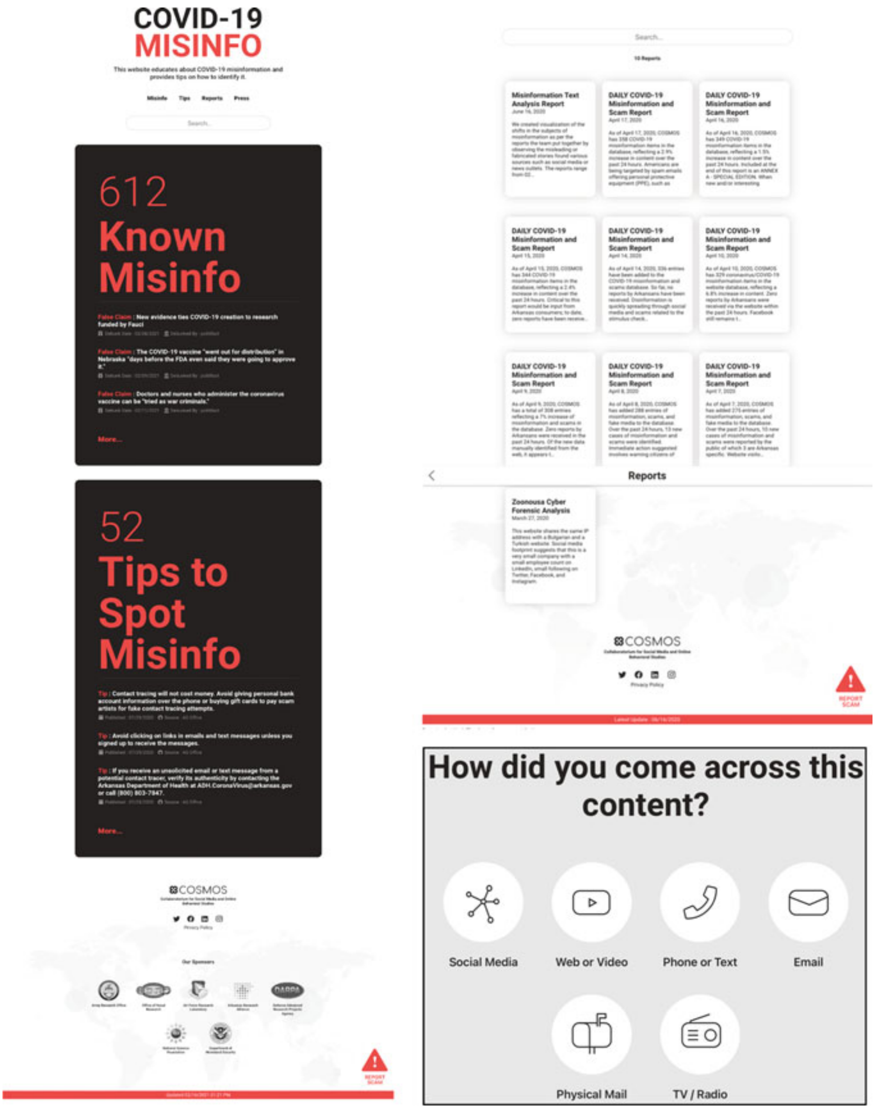


Fig. 3.10 Commenter-Reply Network clusters depicting high echo-chamberness. Square nodes are commenters and circular nodes are repliers. Colors denote the various clusters based on their modularity class

3.5.5 COVID-19 Misinformation Tracker Website

At the request of the Arkansas Attorney General's office, we created a public-facing website (cosmos.ualr.edu/covid-19) that exhibits a smart-curated dataset of known COVID-19 misinformation items and scams in order to reach users in the state and nationwide. Our website also allows users to report additional misinformation items that they encounter, which are then manually validated by our team. This work allows us to coordinate with the Arkansas Attorney General's office to quickly compose and execute alerts to the public. Our COVID-19 Misinformation website and subsequent reporting process resonated well with the Arkansas Attorney General's office and had measurable implications for society. For example, they acted on our numerous suggestions/recommendations for creating communications that they disseminated to the public in order to increase the reach throughout Arkansas, especially in the areas that we showed were not being adequately represented in the website reach statistics (based on user geolocation). These reports are made available to the public via the website (see Fig. 3.11).

The website currently hosts the database of known COVID-19 related misinformation and scams. It is deployed and being used by the Arkansas Attorney General's office. So far, we have documented 612 cases that we identified from numerous



COVID-19
MISINFO

This website educates about COVID-19 misinformation and provides tips on how to identify it.

612
Known Misinfo

False Claim : New evidence ties COVID-19 creation to research funded by Fauci

False Claim : The COVID-19 vaccine "went out for distribution" in Nebraska "days before the FDA even said they were going to approve it."

False Claim : Doctors and nurses who administer the coronavirus vaccine can be "tried as war criminals."

More...

52
Tips to Spot Misinfo

Tip : Contact tracing will not cost money. Avoid giving personal bank account information over the phone or buying gift cards to pay scam artists for fake contact tracing attempts.

Tip : Avoid clicking on links in emails and text messages unless you signed up to receive the messages.

Tip : If you receive an unsolicited email or text message from a potential contact tracer, verify its authenticity by contacting the Arkansas Department of Health at ADH.CoronaVirus@arkansas.gov or call (800) 803-7847.

More...

COSMOS
Collaboratorium for Social Media and Online Behavioral Studies
Privacy Policy
Our Sponsors
Army Research Office
Office of Naval Research
Air Force Research Laboratory
Arkansas Research Alliance
Defense Advanced Research Projects Agency
National Science Foundation
Department of Homeland Security
REPORT SCAM
Search...
10 Reports

Misinformation Text Analysis Report
June 16, 2020
We created visualization of the shifts in the subjects of misinformation as per the reports the team put together by observing the misleading or fabricated stories found various sources such as social media or news outlets. The reports range from 02...

DAILY COVID-19 Misinformation and Scam Report
April 17, 2020
As of April 17, 2020, COSMOS has 358 COVID-19 misinformation items in the database, reflecting a 2.9% increase in content over the past 24 hours. Americans are being targeted by spam emails offering personal protective equipment (PPE), such as

DAILY COVID-19 Misinformation and Scam Report
April 16, 2020
As of April 16, 2020, COSMOS has 349 COVID-19 misinformation items in the database, reflecting a 1.5% increase in content over the past 24 hours. Included at the end of this report is an ANNEX A - SPECIAL EDITION. When new and/or interesting

DAILY COVID-19 Misinformation and Scam Report
April 15, 2020
As of April 15, 2020, COSMOS has 344 COVID-19 misinformation items in the database, reflecting a 2.4% increase in content over the past 24 hours. Critical to this report would be input from Arkansas consumers; to date, zero reports have been receive...

DAILY COVID-19 Misinformation and Scam Report
April 14, 2020
As of April 14, 2020, 336 entries have been added to the COVID-19 misinformation and scams database. So far, no reports by Arkansans have been received. Disinformation is quickly spreading through social media and scams related to the stimulus check...

DAILY COVID-19 Misinformation and Scam Report
April 10, 2020
As of April 10, 2020, COSMOS has 329 coronavirus/COVID-19 misinformation items in the website database, reflecting a 6.8% increase in content. Zero reports by Arkansans were received via the website within the past 24 hours. Facebook still remains t...

DAILY COVID-19 Misinformation and Scam Report
April 9, 2020
As of April 9, 2020, COSMOS has a total of 308 entries reflecting a 7% increase of misinformation and scams in the database. Zero reports by Arkansans were received in the past 24 hours. Of the new data manually identified from the web, it appears t...

DAILY COVID-19 Misinformation and Scam Report
April 8, 2020
As of April 8, 2020, COSMOS has added 288 entries of misinformation, scams, and fake media to the database. Over the past 24 hours, 13 new cases of misinformation and scams were identified. Immediate action suggested involves warning citizens of

DAILY COVID-19 Misinformation and Scam Report
April 7, 2020
As of April 7, 2020, COSMOS has added 275 entries of misinformation, scams, and fake media to the database. Over the past 24 hours, 10 new cases of misinformation and scams were reported by the public of which 3 are Arkansas specific. Website visits...

Reports

Zoonousa Cyber Forensic Analysis
March 27, 2020
This website shares the same IP address with a Bulgarian and a Turkish website. Social media footprint suggests that this is a very small company with a small employee count on LinkedIn, small following on Twitter, Facebook, and Instagram.

COSMOS
Collaboratorium for Social Media and Online Behavioral Studies
Privacy Policy
REPORT SCAM
Latest Update: 06/16/2020

How did you come across this content?

Fig. 3.11 COVID-19 Misinformation Tracker Website Displaying Known Cases, Tips, and Public Reports. Website also allows users to report any cases that is not documented in our database. These cases are investigated before adding to our database

sources (social media—Facebook, YouTube, Twitter, blogs, fake websites, robocalls, text/SMS, WhatsApp, Telegram, and an array of such apps) and from over a dozen countries. We update the database everyday with newly detected cases. Further, we go through identified cases and prepare a list of common telltale signs to detect whether it is a fake website or not. We believe in educating people so that they are self-reliant because we might not be able to detect all possible cases of misinformation. For that reason, we also provide a way for people to submit cases of misinformation that we have not captured in our database. Furthermore, this is not a one-time process. We repeat it every day. So that we can capture as many new cases as possible and expose them. We are also conducting user testing to improve usability of the website. We want to make sure that the website is as easy to use as possible because we realize that our web visitors could be from a wide age group with a varied technical literacy background using different devices. We will expand the system's capability to include tracking features that will help monitor misinformation campaigns as they gain power in real-time. There are several data attributes that are continually documented in our COVID-19 Misinformation website database (see Table 3.1).

As previously mentioned, our COVID-19 Misinformation website and subsequent reporting process resonated well with the Arkansas Attorney General's office and had measurable implications for society. For example, they took immediate action on our numerous recommendations for creating communications that they disseminated to the public in order to increase the reach throughout Arkansas especially in the areas that we showed were not being adequately represented in the website reach (see Fig. 3.12). Although this increase in communications reach may seem small at first glance, it gives evidence that when researchers coordinate with policymakers it can make a difference, especially when that coordination remains an ongoing process.

Table 3.1 COVID-19 misinformation tracker data attributes and description

Data attributes	Description
Title	The general title of the false claim or false narrative
Summary	An extended summary of the false claim or false narrative and includes truths and/or the available evidence that disproves the claims
Disproof	URL link to the article that provides the available evidence that disproves the false claims or false narratives
Date of debunking	Date of the article that provides the available evidence that disproves the false claims or false narratives
Misinformation source	This can have multiple entries, and is the URL, Twitter handle, Facebook username, or other source of the false claim or false narrative
Date of publication	Date of the source of an instance of the false claim or false narrative that is being spread
Country	Used to record the country that is being (or has been) affected by the source of a false claim or false narrative
Keywords	Keywords that are being used to propagate the spread of the false claim or false narrative
Outlet	URL domain, publication, or author (etc.) of a who has been found so far to be propagating the false claim or false narrative

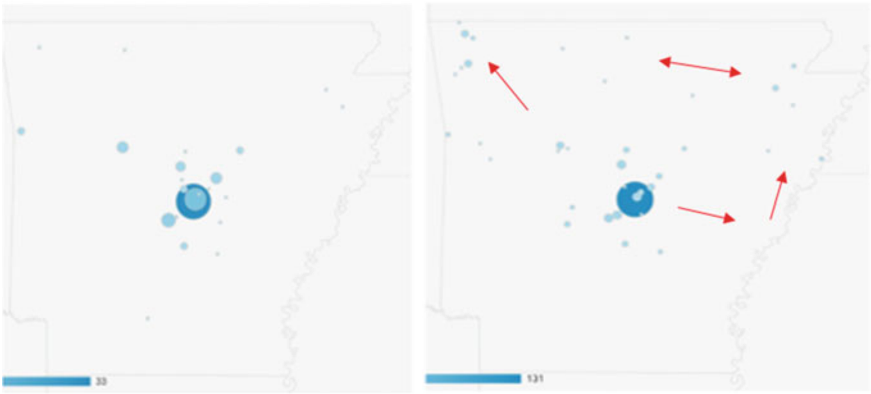


Fig. 3.12 Expansion of the reach of our COVID-19 Misinformation website throughout regions of the state of Arkansas from May 5, 2020 (left) to Jan 21, 2021 (right). Increased penetration into rural areas of Arkansas as annotated by the arrows in the figure on right demonstrates the website's reach and effectiveness of communications

3.6 Conclusion and Future Research Directions

The key takeaways from this chapter include:

1. The misinformation themes of COVID-19 have changed over time since the beginning of the public announcement of the global pandemic. Our research provides evidence that consumer discussions of these themes parallel mainstream media coverage surrounding prominent events. It is important for policymakers to continually utilize their information dissemination resources to reach out to the public in order to provide corrective information.
2. The dissemination of misinformation can be modeled much like that of the spreading of a disease. Such model(s) would help strategize public-health policymaking and crisis communication during pandemics. Information consumers are considered "infectious" when they spread misinformation on social media. Other users become susceptible when they come in contact with that misinformation, and then become "infected" when they decide to continue to spread the misinformation.
3. Our study on connective action demonstrates a way to assess the power of digital communities in mobilizing and coordinating protests. Disinformation campaigns often lead to physical protests. We discuss our research on this method and highlight a case study from the COVID-19 anti-lockdown protests in Michigan showing that connective action from these communities is more cohesive and thus more successful in their protest efforts. Identifying these connective action events before reaching critical mass will allow policymakers to engage in appropriate and timely countermeasures.

4. Toxicity on online social networks has the power to not only cause disruption in public discourse on social media platforms but can also cause the formation of segregated and polarized communities within those networks. We have provided computational evidence of how such toxicity can be identified, but also how removing highly toxic users from a network improves the state of online discourse by reducing hate speech and healing fractured digital communities. Although this is intuitive, it is up to the policymakers within each online social network to act on this evidence in order to continue to be dedicated to providing a healthy platform for their users.
5. When researchers actively coordinate with policymakers, the problem of misinformation can be effectively communicated to educate information consumers. Evidence of this is given by the example of our work with the Arkansas Attorney General's Office, wherein they had measurable results in increasing their communications reach within the state regarding the existence of COVID-19 misinformation campaigns and scams that affect their citizens.

Although the use of misinformation has a long history, the problem can be overcome with continued collaborative efforts by researchers and policymakers. The most prominent concept to glean from this chapter is the importance of bridging the gap between research and policy. For example, one of our most successful endeavors is our work for the Arkansas Attorney General's office, wherein we were able to successfully aid in the creation and dissemination of public communications regarding the existence of various COVID-19 misinformation campaigns. Our COVID-19 Misinformation tracker website and subsequent reporting process had measurable implications for society. When the Arkansas Attorney General's office acted on our suggestions and recommendations for creating communications and disseminating them to the public, they were able to increase their reach throughout Arkansas especially in rural areas. Methods such as the ones utilized in our studies can be useful when incorporated into the moderation processes of online social networks and into the strategic planning efforts of public-health policymakers. Our research methodologies combine analysis from publicly available datasets and data collected in this effort.

We are observing a real-time shift in behaviors of the online social media information environment. Due to recent shutdowns of many social media accounts, fringe users are leaving popular platforms for other more secure and less restrictive or less moderated modes of communication. What does this mean for the social scientist and those that perform social network analysis? It means that we have to keep up with this changing online information environment and adapt our data collection and analysis strategies to stay abreast with cyber behavior evolution. It is important to continually identify and track the trajectory of these behaviors. With continual advancements in big data analysis techniques, researchers and policymakers can come together to better detect, mitigate, and reduce misinformation that is plaguing online social networks. Ongoing research within this realm includes improving automated misinformation detection techniques, understanding the dynamics of various cross-platform information positioning tactics (especially on emerging social media platforms), modeling these behaviors to better enable prediction of propagation actors and tactics, and continually fine-tuning best practices for communicating research to policymakers and taking effective action to better society.

Acknowledgements This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

Agarwal, N., Al-khateeb, S., Galeano, R., & Goolsby, R. (2017). Examining the use of botnets and their evolution in propaganda dissemination. *NATO Journal for Defense Strategic Communications, 2*, 87–112.

Agarwal, N., & Bandeli, K. K. (2020). Examining strategic integration of social media platforms in disinformation campaign coordination. *NATO Journal for Defense Strategic Communications, 4*(1), 173, 2018.

Agarwal, N., Lim, M., & Wigand, R.T. (2012). Raising and rising voices in social media: A novel methodological approach in studying cyber-collective movements. *Business & Information Systems Engineering (BISE) Special Issue on IS and Culture,* Dorothy Leidner (Ed.), WIRTSCHAFTSINFORMATIK. 2012. https://doi.org/10.1007/s12599-012-0210-z

Alassad, M., Hussain, N., & Agarwal, N. (2020). *Systems thinking and modeling in social networks: A case study of coronavirus conspiracy theories challenges in Twitter network*. Presented at the SBP-BRiMS 2020, October 2020.

Al-khateeb, S., & Agarwal, N. (2019). Deviance in social media and social cyber forensics: Uncovering hidden relations using open source information (OSINF). *Springer Briefs in Cybersecurity.* Springer, 2019. ISBN: 978-3-030-13689-5.

Al-khateeb, S., Conlan, K. J., S., Nitin, A., Ibrahim, B., & Frank, B. (2016). Exploring Deviant Hacker Networks (DHM) on social media platforms. *Journal of Digital Forensics, Security and Law, 11*. https://doi.org/10.15394/jdfsl.2016.1375.

Allyn, B. (2020, June 16). *Study exposes Russia disinformation campaign that operated in the shadows for 6 years.* NPR. https://www.npr.org/2020/06/16/878169027/study-exposes-russia-disinformation-campaign-that-operated-in-the-shadows-for-6-. Last accessed 28 Feb 2021.

Ananyev, M., & Sobolev, A. (2017, April). *Fantastic beasts and whether they matter: Do internet 'trolls' influence political conversations in Russia?* Presented at Midwest Political Science Association, Chicago, IL.

Anderson, J., & Moody, K. (2020). Stigmas and filter bubbles: How the Russian misinformation campaign in the 2016 US presidential campaign links to stigma communication. *Iowa Journal of Communication, 52*(2), 31–47.

Badawy, A., Ferrara, E., & Lerman, K. (2018, August). *Analyzing the digital traces of political manipulation: The 2016 Russian interference twitter campaign*. In 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM) (pp. 258–265). IEEE.

Banerjee, D., & Rao, T. S. (2020). Psychology of misinformation and the media: Insights from the COVID-19 pandemic. *Indian Journal of Social Psychiatry, 36*(5), 131.

Bao, H., Cao, B., Xiong, Y., & Tang, W. (2020). Digital media's role in the COVID-19 pandemic. *Journal of Medical Internet Research mHealth and uHealth*, *8*(9), e20156.

Barthel, M., & Mitchell, A. (2017, May 10). Americans' attitudes about the news media deeply divided along partisan lines. *Pew Research Center's Journalism Project*. https://www.journalism.org/2017/05/10/americans-attitudes-about-the-news-media-deeply-divided-along-partisan-lines/. Accessed 13 Nov. 2019.

Bettencourt, L. M., Cintron-Arias, A., Kaiser, D. I., & Castillo-Chavez, C. (2006). The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica a: Statistical Mechanics and Its Applications, 364*, 513–536.

Bond, S. (2021, February 8). *Facebook widens ban on COVID-19 vaccine misinformation in push to boost confidence*. NPR. https://www.npr.org/2021/02/08/965390755/facebook-widens-ban-on-covid-19-vaccine-misinformation-in-push-to-boost-confiden. Last accessed 28 Feb 2021.

Burkhardt, J. M. (2017). History of fake news. *Library Technology Reports, 53*(8), 5–9.

Burton, A., & Koehorst, D. (2020). Research note: The spread of political misinformation on online subcultural platform. *Harvard Kennedy School, Misinformation Review*. https://misinforeview.hks.harvard.edu/article/research-note-the-spread-of-political-misinformation-on-online-subcultural-platforms/. Last accessed 28 Feb 2021.

Calabresi, M. (2017). *Inside Russia's social media war on America*. https://time.com/4783932/inside-russia-social-media-war-america/. Last accessed 23 Feb 2021.

Center for Strategic & International Studies. (2020 September 23). *Countering Russian disinformation*. https://www.csis.org/blogs/post-soviet-post/countering-russian-disinformation. Last accessed 28 Feb 2021.

Cha, M., Gao, W., & Li, C. (2020). Detecting fake news in social media: An Asia-Pacific perspective. *Communications of the Association for Computing Machinery, 63*(4), 68–71. https://doi.org/10.1145/3378422

Clark, M. (2020 September 19). Misinformation poses 'severe test' in Asia Pacific. *Food Safety News*. https://www.foodsafetynews.com/2020/09/misinformation-poses-severe-test-in-asia-pacific/. Last accessed 28 Feb 2021.

Chou, W. Y., Gaysynsky, A., & Cappella, J. N. (2020). *Where we go from here: Health misinformation on social media*.

Commission on Security and Cooperation in Europe. (2017, September 14). *The scourge of Russian disinformation*. https://www.csce.gov/international-impact/events/scourge-russian-disinformation. Last accessed 28 Feb 2021.

Commonsense.org. (2017). *Fake news: Historical timeline*. https://www.commonsense.org/education/sites/default/files/tlr-asset/newsmedialit_fakenewstimeline_8.5x11.pdf. Last accessed 28 Feb 2021.

Compartmental models in epidemiology. (2021, February 21). In *Wikipedia*. https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

Fernandez, B. (2019). Fake news in Asian politics. *Global Risk Insights*, 1.

Fernandez, M. (2021, February 6). Anti-vaccine activists emboldened in California. *The New York Times*. https://www.nytimes.com/2021/02/06/us/california-covid-vaccine.html. Last accessed 28 Feb 2021.

Food and Agricultural Organization of the United Nations. (2020, November 17). *Communicating food safety in the era of COVID-19—Earning consumer trust*. http://www.fao.org/asiapacific/news/detail-events/en/c/1330594/. Last accessed 28 February 2021.

Galeano, K., Galeano, R., & Agarwal, N. (2020). An evolving (dis)information environment—How an engaging audience can spread narratives and shape perception: A trident juncture 2018 case study. 253–265. https://doi.org/10.1007/978-3-030-42699-6_13

Galeano, K. K., Galeano, L. R., Mead, E., Spann, B., Kready, J., & Agarwal, N. (2019). *The role of YouTube during the 2019 Canadian Federal election: A multi-method analysis of online discourse and information actors.* Queen's University.

Grant, M. (2004). *Greek and Roman historians: Information and misinformation.* Routledge.

Horton, R. (2020). Offline: Managing the COVID-19 vaccine infodemic. *Lancet (london, England), 396*(10261), 1474.

Jerit, J., & Zhao, Y. (2020). Political misinformation. *Annual Review of Political Science, 23*(1), 77–94.

Kearney, M., Chiang, S., & Massey, P. (2020). The Twitter origins and evolution of the COVID-19 "plandemic" conspiracy theory. *Harvard Kennedy School, Misinformation Review* (9). https://misinforeview.hks.harvard.edu/article/the-twitter-origins-and-evolution-of-the-covid-19-plandemic-conspiracy-theory/. Last accessed 28 Feb 2021.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115*(772), 700–721.

Lee, J. (2021, January 11). The utter familiarity of even the strangest vaccine conspiracy theories. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2021/01/familiarity-strangest-vaccine-conspiracy-theories/617572/. Last accessed 28 Feb 2021.

Loomba, S., de Figueiredo, A., Piatek, S., de Graaf, K., & Larson, H. J. (2020). *Measuring the impact of exposure to COVID-19 vaccine misinformation on vaccine intent in the UK and US*. medRxiv.

Maleki, M., Mead, E., Arani, M. & Agarwal, N. (2021). *Using an epidemiological model to study the spread of misinformation during the black lives matter movement*. International Conference on Fake News, Social Media Manipulation and Misinformation (ICFNSMMM). Athens, Greece.

Marcoux, T., Mead, E., & Agarwal, N. (2020). *The ebb and flow of the COVID-19 misinformation themes.* Presented at the SBP-BRiMS.

Mason, L. E., Krutka, D., & Stoddard, J. (2018). Media literacy, democracy, and the challenge of fake news. *Journal of Media Literacy Education, 10*(2), 1–10.

McCourt School of Public Policy. (2020, September 24). *Misinformation and the impact of social media in elections.* Georgetown University. https://mccourt.georgetown.edu/news/misinformation-and-the-role-of-social-media-in-elections/. Last accessed 28 Feb 2021.

Mihailidis, P., & Viotty, S. (2017). Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in "post-fact" society. *American Behavioral Scientist, 61*(4), 441–454.

Miller, J. (2019, June 10). *Democracy and misinformation: The Cold War and today.* The Newsmagazine of the American Historical Association: Perspectives on History. URL: https://www.historians.org/publications-and-directories/perspectives-on-history/summer-2019/democracy-and-misinformation. Last accessed 28 Feb 2021.

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014*). Information network or social network? The structure of the Twitter follow graph.* Proceedings of the 23rd International Conference on World Wide Web, 493–498. https://doi.org/10.1145/2567948.2576939

Obadimu, A., Khaund, T., Mead, E., Marcoux, T. & Agarwal, N. (2021). Developing a socio-computational approach to examine toxicity propagation and regulation in COVID-19 discourse on YouTube. *Information Processing and Management Special issue on Dis/Misinformation Mining from Social Media, 58*(5). https://doi.org/10.1016/j.ipm.2021.102660

Palatino, M. (2019, September 26). *Combatting disinformation in Asia Pacific: Intended—And unintended—Consequences.* IFEX. https://ifex.org/combatting-disinformation-in-asia-pacific-intended-and-unintended-consequences/. Last accessed 28 Feb 2021.

Pavlopoulos, J., Thain, N., Dixon, L. & Androutsopoulos, I. (2019). *ConvAI at SemEval-2019Task 6: Offensive language identification and categorization with perspective and BERT*. Proceedings of the 13th International Workshop on Semantic Evaluation, 571–576. https://doi.org/10.18653/v1/S19-2102

Reiss, D. (2021, January 20). *COVID-19 vaccine misinformation and the anti-vaccine movement*. Harvard Law, Petrie-Flom Center. https://blog.petrieflom.law.harvard.edu/2021/01/20/covid-19-vaccine-misinformation-anti-vaccine-movement/. Last accessed 28 Feb 2021.

Salzberg, S. (2020, February 3). How anti-vax activists use conspiracy theories to spread fear of vaccines. *Forbes*. https://www.forbes.com/sites/stevensalzberg/2020/02/03/how-the-anti-vaccine-cult-spreads-its-message/?sh=6318da5c2036. Last accessed 28 Feb 2021.

Sanovich, S. (2017). *Computational propaganda in Russia: The origins of digital misinformation.* S. Woolley & P. N. Howard (Eds.), Working Paper 2017.3. Project on Computational Propaganda. comprop.oii.ox.ac.uk. 32 pp.

Sanovich, S., Stukal, D., & Tucker, J. A. (2018). Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia. *Comparative Political Studies, 50*(3), 435–482.

Segerberg, A., & Bennett, L. (2013). *The logic of connective action: Digital media and the personalization of contentious politics.* Cambridge University Press. https://doi.org/10.1017/CBO9781139198752

Seitz, A. (2021, February 24). *Fake accounts gain traction as they praise China, mock US*. AP NEWS. https://apnews.com/article/media-social-media-coronavirus-pandemic-covid-19-pandemic-china-7339598fed868fcfe109999bf071a77c. Last accessed 28 Feb 2021.

Shahi, K.G., Dirkson, A., & Majchrzak, T. A. (2020). *An exploratory study of COVID-19 misinformation on Twitter*. arXiv e-prints, arXiv-2005.

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. 745–750. https://doi.org/10.1145/2872518.2890098

Shu, K., Sliva, A., Wang, S., Tand, J., & Liu, H. (2017). Fake news detection: Network data from social media used to predict fakes. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.

Søe, S. O. (2018). Algorithmic detection of misinformation and disinformation: Gricean perspectives. *Journal of Documentation, 74*(2), 309–332.

Soll, J. (2016, December 18). The long and brutal history of fake news. *Politico Magazine.* https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/. Last accessed 28 Feb 2021.

Spann, B., Agarwal, N., Johnson, O., & Mead, E. (2020, October 18–21). *Modeling protester orchestration through connective action: A COVID-19 lockdown protest case study.* 2020 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction, and Behavior Representation in Modeling and Simulation (SBP-BRiMS 2020). Washington, DC.

Starmans, B. (2018). *10 examples of fake news from history*. The Social Historian. https://www.thesocialhistorian.com/fake-news/. Last accessed 28 Feb 2021.

Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: Literature review and synthesis. *Annals of the International Communication Association, 44*(2), 157–173.

Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism, 7*(6), 802–823.

Watts, C. (2018). Messing with the enemy: Surviving in a social media world of hackers terrorists, Russians, and fake news. *Harper Business.*

Weitz, R. (2020). ICDS Diplomaatia magazine, November 13, 2020. Assessing the Russian Disinformation Campaign During COVID-19. https://icds.ee/en/assessing-the-russian-disinformation-campaign-during-covid-19/. Last accessed 22 Feb 2021.

Wendling, M. (2018). *The (almost) complete history of 'fake news'*. BBC. https://www.bbc.com/news/blogs-trending-42724320. Last accessed 28 Feb 2021.

Wilson, T., & Starbird, K. (2020). Cross-platform disinformation campaigns: Lessons learned and next steps. *Harvard Kennedy School: Misinformation Review, 1*(1), 2020.

Zhang, X., & Ghorbani, A. A. (2019, August). *An overview of online fake news: Characterization, detection, and discussion.* Information Processing and Management (2018), 102025. https://doi.org/10.1016/j.ipm.2019.03.004

Chapter 4
COVID-19 Disinformation, Misinformation and Malinformation During the Pandemic Infodemic: A View from the United Kingdom

Neil Verrall

Abstract The global pandemic has highlighted that despite the vast medical, economic and social mitigations being rightly focused on its impact, the use of disinformation, misinformation and malinformation (DMM) was conducted by a range of threat actors in order to continue their malign activities in support of their geopolitical, strategic and operational objectives. This chapter describes DMM during the pandemic, as well as commentary on the need to start understanding how DMM has evolved into a capability or system, rather than a discrete, isolated tactic; and how DMM is likely to evolve given anticipated future developments within society and technology, which should be of interest to the defence and security domain.

4.1 Introduction

The coronavirus (COVID-19) pandemic demonstrated that old (unattributed) maxim—the enemy never sleeps. In the days and weeks after the outbreak started hitting the global media apparatus, news began to appear about COVID-related disinformation and misinformation. This was despite two appeals by the United Nations' Secretary-General António Guterres in late March and early April 2020 (United Nations, 2020a, 2020b),where he called for "good will from old adversaries" and that global armed actors put down their weapons [physical and digital] in order to help vulnerable populations cope and mitigate the pandemic. This did not happen, and rather quickly reports of disinformation and misinformation began to circulate among the world's media and the fact-checking industry. Disinformation is now considered *de rigueur* for major events; for example, it was employed to hamper the post-event analysis for the shooting down of Malaysia Airlines Flight 17 in July 2014

N. Verrall (✉)
Defence Science and Technology Laboratory, Salisbury, UK
e-mail: ngverrall@dstl.gov.uk

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_4

(EUvsDisinfo, 2019), the Novichok poisonings in the UK in March 2018 (Gunter & Robinson, 2018) and is a part of most (if not all) national elections where hostile state actors and non-state actors have regional and/or geopolitical interests. Therefore, it seems prudent to understand the pattern of activity for disinformation during the recent COVID-19 pandemic.

4.2 The Nature of Disinformation

The pandemic demonstrated the inevitable presence of both disinformation and misinformation—or what the Director-General of the World Health Organisation (WHO) referred to at the 15 February 2020 Munich Security Conference as an *infodemic*—"we're not just fighting an epidemic; we're fighting an Infodemic" (WHO, 2020a). The WHO defines an Infodemic as an "overabundance of information—some accurate and some not—that occurs during an epidemic. It can lead to confusion and ultimately mistrust in governments and public health response" (WHO, 2020b). According to the UK's Online Harms White Paper (HM Government, 2019a) disinformation is defined as "information which is created or disseminated with the deliberate intent to mislead; this could be to cause harm, or for personal, political or financial gain", whilst misinformation is the "inadvertent sharing of false information." Increasingly, the term *malinformation* is creeping into this domain. Whereas disinformation and misinformation focus explicitly on false information (whether deliberately created or inadvertently shared), malinformation deals with leaks, harassment and hate speech, which possesses *intent to harm*, but may not be clearly false or fake (Wardle, 2019). All three terms contribute to what could usefully be referred to as *information disorder* (Wardle, 2019), within a society that is fast becoming an information civilization (Zuboff, 2015).

The use of deliberate false information for military or political purposes is immemorial, and disinformation has quite rightly been referred to as "an old story, fuelled by new technology" (Posetti et al., 2018). Today it presents itself in different forms, different formats and from different sources; for example, from state actors, non-state actors, proxies (which includes pro-state/pro-cause patriotic trolls), the criminal community (incl. scam artists and fraudsters), as well as malevolent trolls with nothing better to do. In order to better understand the disinformation, misinformation and malinformation (DMM) strategies of adversaries, numerous researchers have posited several models and theories (Table 4.1), as well as identifying various tradecraft and techniques used to create and deliver DMM (Table 4.2).

Table 4.1 Models and theories of DMM

Model	Characteristics
Verification Handbook	Confirm authenticity (provenance); confirm source; confirm location; confirm date of event (Silverman, 2014)
Firehose of Falsehood	High-volume and multichannel; rapid, continuous and repetitive; lacking commitment to objective reality; and lacking commitment to consistency (Paul & Matthews, 2016)
4D model	Dismiss, Distort, Distract and Dismay (White, 2016)
Digital containment	1. Strategic asymmetry; tactical flexibility; plausible deniability 2. Unsourced or falsified claims; non-credible claims with sources; claims based on earlier unsourced or non-credible claims; conspiracy theories (Bjola & Pamment, 2016)
F.I.R.S.T	Fabrication; Identity, Rhetoric, Symbolism, Technology (Government Communication Service [GCS], 2019)
Seven categories of information disorder	Satire or parody; misleading content; imposter content; fabricated content; false connection; false content; manipulated content (Wardle, 2019)
Cardiff Seven factors	Seeding; denial of credibility; event ghosting; emulsifying; infiltrating and inciting; spoofing; and 'truthing' through social proofing (Innes, 2020)
Ecosystem Pillars	Official government communications; state-funded global messaging; cultivation of proxy sources; weaponisation of social media; cyber-enabled disinformation (Global Engagement Center [GEC], 2020a)
Spotting false behaviour	Account automation; Profile; Posts; Point of View (Miller & Colliver, 2020)

4.3 The Unfolding Infodemic

One thing is clear, the pandemic was subject to an awful lot of DMM. The Poynter Institute in the U.S. had previously set-up the International Fact Checking Network (IFCN), with dozens of fact-checking organisations in over 70 countries and 40 languages. In order to address the inevitable disinformation around the pandemic the Poynter Institute set-up the Coronavirus Facts Alliance Database. A week after it publically launched (the last week of March 2020) the number of facts checked

Table 4.2 Techniques and tradecraft of DMM

Engagement techniques[a]	Ad hominem; astroturfing; bandwagoning; butler lies; card stacking; cheerleading; conspiracies; false dilemma; flooding; laundering; gish gallop; transfer; strawman; misappropriation; ping pong; point and shriek; Potemkin village; raiding; satire, parody and ridiculing; shilling; sockpuppets; spiral of silence; symbolic action; tainting; totum pro parte; trolling; whataboutism; wolf cries wolf; woozle effect
Tradecraft	• **Accounts** (false accounts, aged accounts, hijack accounts) • **Attribution** (conceal identity, conceal location [IP], post and delete [source]) • **Personas** (voice, speech, text, gesture [e.g. deepfakes] and other images (e.g. photos, symbols, icons, etc.) • **Content** (mirror websites, false verification badges, imposter content, fake or edited documents, dark ads, tainted leak [false within truth], compelling infographics, buy social media metrics) • **Behaviours** (befriending, lurking, phishing, trolling) • **Networks** (mutual following, amplify, botnets, nexus of agents of influence)

[a] Abridged from GCS (2019) and Center for European Policy Analysis (2020)

Table 4.3 Key themes (and examples) of COVID-19 disinformation

Causes (the origin story)
Examples included the U.S releasing the virus (CIA or soldier); the Chinese releasing the virus; the virus is an excuse to create new bioweapons; created to interfere with US elections; and as a way of creating money for elites; ‘Big Pharma’; a previous patent for a US vaccine; fragments from a comet; that it was bioengineered by either the USA or China; it started in Italy
The Spread
Examples included 5G signal towers; pets; bananas; lack of medical insurance; variations of coffins and corpses stories; a raft of contexts linking intentional infections conducted by certain countries; and many countries having their own commentary on immigrants and ‘foreigners’
Prevention and Cures
Examples included the use of saline; coffee; colloidal silver; zinc; hydroxychloroquine; chloroquine; salt; sunlight; alcohol; vitamin C; alkaline foods; tonic water; ultraviolet lamps; eucalyptus oil vapour; mineral solutions; baking soda; gargling bleach; nasal washing; ayurvedic medicine; vegan diets; cold showers; steam; carbon pills; homeopathy; cannabis; cocaine; sex

regarding pandemic-specific disinformation went from approximately 800 to over 1,500. On International Fact Checking Day (2 April 2020) the number exceeded 3,000 facts checked, which was approaching 4,000 by end of April. It then exceeded 5,000 by mid-May and 7,000 by the end of July. The long tail continued and reached over 9,000 facts checked by December 2020. Within this deluge three dominant themes emerged: (1) the supposed causes of the virus (2) how it was supposedly spread, and (3) supposed prevention and cures (Table 4.3).

4.4 Russia

The terms 'Russia' and 'disinformation' (*dezinformatsiya* [дезинформация]) seem synonymous—both historically and contemporarily. It is not definitively known when the term entered into the Russian lexicon, although it has been attributed to the forerunner of the KGB,[1] which was the State Political Directorate,[2] and the establishment of a Russian state disinformation office in 1923 (Manning & Romerstein, 2004). The term does, however, also appear in an Armenian text on the Middle East at the turn of the twentieth century (Ghazari-Sargsyan et al., 1900). Nevertheless, there is certainly increased evidence of its use within Russian military publications in the 1920s and 30s, in publications such as *Roman-Gazeta* (Roman-Gazeta, 1927) and *War and Revolution* (Pochter, 1932).

Disinformation was embraced during the Great Patriotic War (1941–1945), especially alongside *maskirovka* [маскировка], which were camouflage and concealment activities at the military tactical and operational levels; therefore, false information supported false physical signatures within military stratagem (Matsulenko, 1975). *Dezinformatsiya* was first formally recorded by the Allies in their 1945 Russian military dictionary (U.S. War Department, 1945), but it was during the Cold War (1947–1991) that disinformation was used as a key activity against the West (Bittman, 1985; CIA, 1986; Kux, 1985; Schultz & Godson, 1984; US Government, 1987). In particular, Operation INFEKTION provides a good example of a Soviet health-related disinformation campaign in the 1980s, which exploited the context of the HIV/AIDS issue at the time (Boghardt, 2009). Operation INFEKTION also helps to demonstrate that the use of disinformation within a health or medical-related context is not a new attack vector; and this was demonstrated again by the Russia during the 2013 Ebloa outbreak.

Disinformation is identified as an activity within several models and theories of Russian deception and maskirovka (see Korolev et al., 2015; Koziratsky et al., 2015; Marushchenko & Klevtsova, 2014; Matveevsky, 2017; Orlyansky & Kuznetsov, 2013); and Serov (2011, 2019) describes disinformation as "the provision of deliberately false information to the enemy for more effective conduct of hostilities, checking for information leakage and the direction of its leakage, as well as the process of manipulating it, misleading someone by providing incomplete or complete, but no longer necessary information, but also distortion of its part or presentation of tendentious and deliberately false information."

During the early stages of the pandemic, between mid-January to mid-March 2020, the European Union External Action Service's (EEAS) East Strategic Communications Task Force (who focus on pro-Kremlin media outlets) collected over 110 COVID-19-related disinformation cases in their public EUvsDisinformation database (EUvsDisinfo, 2020a). This exceeded 300 by mid-April and 500 by mid-May (EUvsDisinfo, 2020b).

[1] KGB (Komitet Gosudarstvennoy Bezopasnosti [Комитет государственной безопасности]).

[2] SPD or GPU (Gosudarstvennoe politicheskoe upravlenie [Государственное политическое управление]).

In March 2020 the U.S. GEC published analysis showing how a surge in disinformation activity by Russian state-linked false accounts, personas and proxies began on 20 January 2020 (GEC, 2020b). This was subsequently followed-up by the GEC's report that outlined the Russian disinformation ecosystem (GEC, 2020a); within this report was a significant amount on COVID-19 disinformation. Some of the evidence presented showed how pro-Kremlin proxy websites spread disinformation, such as conspiracy theories (e.g. cause of the virus, role of elites), blaming the U.S and the West, and many of the narratives and tropes identified in Table 4.3.

NATO's Defender Europe 20 exercise was also subjected to DMM, whereby NATO was accused of deploying 30,000 troops across Europe who did not have to wear masks, just as the pandemic was spreading across Europe. The exercise was also criticised for the intention to hold a series of free rock concerts in several NATO countries, which was seen as conflicting with physical distancing advice. The European Union and the WHO authorities were accused of turning a blind eye and were subjected to satire whereby the sophists accused them of wearing "the mask on over their eyes, as well as over their mouth and nose" (Dinucci, 2020).

Another example was when the U.S. publication Defense One communicated that the U.S. Air Force had transported half-a-million coronavirus test swabs from Italy to Tennessee in order to meet the growing U.S. homeland demand (Weisgerber, 2020). This was picked-up by pro-Kremlin media and reinterpreted as the U.S. removing vital medical supplies and test kits that could have been used to support their NATO ally (Ibragimova, 2020)—the subtlety being the interpretative slant on whether the U.S were providing much needed supplies to their own citizens or whether they were denying help to their NATO ally, i.e. the use of the 'whataboutism' technique (Table 4.2). Additionally, the word 'swab' was exchanged for 'kit', thereby giving a false impression of what the equipment actually was (Andriukaitis, 2020). By mid-April NATO had identified five myths that Russian or pro-Russian media had been circulating online: (1) COVID-19 will break-up NATO, (2) NATO is failing to support its allies, (3) COVID-19 is a weapon created by NATO, (4) NATO exercises spread COVID-19, and (5) NATO encourages defence spending at the expense of healthcare (NATO, 2020). At the end of April the NATO Secretary-General accused Moscow and Beijing of spreading DMM throughout the pandemic regarding NATO's response to the situation (Jozwiak, 2020a).

In Bulgaria the chairman of the pro-Russian Revival Party was indicted with spreading false information about COVID-19 that could cause panic; although the rebuttal pointed at the indictment being a politically motivated diversion (Radio Free Europe/Radio Liberty, 2020). North Macedonia joined NATO on 19 March 2020 and was subject to an upsurge in COVID-19 DMM. North Macedonia had expected this because Montenegro had also been on the receiving end from Russian DMM and cyberattacks when it joined NATO in 2017 (Marusic, 2020) and North Macedonia had also been subjected to Russian DMM during their 2018 referendum to change the name of the state (Metodieva, 2018).

In late April the Russian Foreign Ministry made highly misleading comments regarding "explosive" COVID-19 infection rates among U.S. service personnel in Syria (Tass, 2020). This was despite the U.S military having an infection rate of 0.2%

among its 1.3 million military personnel. At the same point in time a well-known journalist and fact checker in Lithuania was subjected to an *imposter content* attack whereby:

> A forged letter from NATO Secretary General Jens Stoltenberg was sent to Lithuanian media, NATO headquarters in Brussels, and other governmental institutions from a spoofed email of the Lithuanian Ministry of Defense (MoD). The letter alleged that NATO is going to withdraw its troops from Lithuania due to the COVID-19 pandemic. The letter was written in poor English, and a version of Stoltenberg's signature that is openly available online matched perfectly with the one used in the letter. (Aleksejeva, 2020a)

In a demonstration of DMM tradecraft (Table 4.2) the forged letter was used as evidence and published on a You Tube channel that had been created on 22 April, but overall the story did not gain traction on social or mainstream media. The journalist in question had previously been subjected to numerous identify thefts, with fake blogs and channels set-up in his name. The technique of imposter content is an increasingly popular way to try and undermine and discredit fact checkers and the fact checking community. In neighbouring Latvia, the fact checking team at Re:Check was targeted by individuals and groups they had previously debunked. The antagonists petitioned Facebook in order to get them to break their partnership with Re:Check. The aim of this *conspiracists unite* tactic is to harass, undermine and discredit the fact checking organisation in order to curtail their analysis or simply to distract their attention from focusing on their tasks of monitoring dubious sources, fact checking and debunking (Aleksejeva, 2020b). The use of these DMM techniques suggests that the fact checking industry must be doing something right if they are being targeted by the sophists.

In Africa, Russia exploited a French *faux pas* when French medical scientists suggested that a future COVID-19 vaccine should first be tested in Africa. The resulting outrage produced hashtags such as #Africansarenotlabrats and #Africansarenotguineapigs resulting in Russian state-sponsored media outlets amplifying the outrage; including Central African-based Facebook group *Jeunesse derrière la Russie* (Youth behind Russia). This should be seen as part of Russia's broader influence operations in Africa (Grossman et al., 2019; Limonier, 2019; Shekhovtsov, 2020).

An additional narrative pushed by adversaries, particularly Russia and China, was the victimisation narrative. The EUvsDisinformation database contains over 600 results for 'Anti-Russian' disinformation stories, with 12 these stories between February and May 2020 exploiting COVID-19 as the pretext for victimisation (Table 4.4). Russia also supported China by portraying the West, NATO and the EU as victimising bullies of China (EUvsDisinfo, 2020c)—a case of my enemy's enemy is my friend—and if not a friend per se then at least a kindred spirit or tolerable neighbour.

Russian disinformation also formed part of their aid and assistance narrative, whereby it was falsely claimed that Poland would not allow Russian aircraft to fly through Polish airspace in order to deliver aid to Italy (EUvsDisinfo, 2020d). This touches on the idea of a "bad Samaritan" (Braw, 2020), which is the juxtaposition of an adversary who also does (or is trying to be seen to be doing) something good

Table 4.4 COVID-19 anti-Russian disinformation stories

• Russia is being falsely accused of waging a disinformation campaign against COVID-19
• Western media propaganda claims that COVID-19 is a project of the Kremlin
• The West is attacking Russia on COVID-19
• NATO waging disinformation campaign against Russia's COVID-19 aid to Italy
• In conditions of an epidemic, the EU continues to promote the concept of the 'Russian Enemy'
• COVID-19 has given the Baltic States a new excuse for Russophobia
• The West plans to attack China financially for the COVID-19 outbreak just as it did to Russia
• EU demonises Russia, attributing all its problems to it
• Amid COVID-19 epidemic, Latvia intensifies neo-Nazism and oppression of Russian-speaking population
• Ukrainian Russophobes behind La Stampa's anti-Russian disinformation campaign
• EU cites zero facts about Russia's alleged disinformation on COVID-19
• Poland continues to tear off Ukraine and Belarus from Russia even during the pandemic

occasionally, i.e. good people can do bad things, and bad people can, sometimes, do good things—it all depends what the motivation is for the behaviour. In the grey zone of belligerent activities that occur below the traditional threshold for war it seems completely legitimate for hostile states to say and do one thing (something good), whilst simultaneously doing something else and something different (something bad). This may be an uncomfortable concept for some people to acknowledge, but we no longer compete in a conflict space that is binary and clearly defined, which is why the literature on the 'grey zone' has emerged (Hoffman, 2018; Mazarr, 2015; Morris et al., 2019), as well as associated DMM and influence-related publications (Verrall & Mason, 2018; Verrall et al., 2019).

Russia also used the pandemic as an opportune moment to conduct wider antagonistic behaviour; for example, they deployed mobile brigades to Georgia's occupied Abkhazia region, barring entry to all foreigners except Russians, including the posting of forces at the only crossing point connecting the region with the rest of Georgia (Agenda, 2020a). This pattern of Russian behaviour is known as 'borderization' (Kakachia, 2018) or 'creeping occupation' (IDFI, 2015) and has been an enduring tactic since the Russo-Georgian War of 2008 (Khan, 2019) Russia's behaviour during the pandemic was recognised by the international community, and several European ambassadors and ministers of foreign affairs commented on the situation, with the Lithuanian Foreign Minister stating that Russia's behaviour was "serving as a smokescreen for further borderization" (Agenda, 2020b).

In summing up the Russian focus, an article published by the Atlantic Council identified a six factor strategy as to why the Kremlin would spread disinformation about the pandemic: (1) to spread anti-west/NATO/US messages to weaken them, (2) to sow chaos and panic, (3) to undermine the target audience's trust in credible sources of information, (4) to undermine trust in objective facts, (5) to spread conspiracies to facilitate acceptance of other conspiracies, (6) assist Russia's ability to identify the channels spreading disinformation (Kalenský, 2020). This aligns with analysis by the GEC (2020a) who stated that Russia had "used the global pandemic as a hook to push longstanding disinformation and propaganda narratives."

4.5 China

China was initially accused of disinformation but it was felt that they were fighting an obvious losing battle, thus, they decided to switch tactics for an alternative narrative strategy, which emphasised their role in containing the pandemic, economic recovery and global engagement (Australian Strategic Policy Initiative [ASPI], 2020). However, numerous examples of Chinese DMM did emerge; for example, Taiwan was subject to a series of false messages in late February that Taiwan's Investigation Bureau attributed to China (Thomas & Zhang, 2020). The focus of this DMM was on Taiwan's inability to control the virus and that the government was hiding the truth. Analysis by the Taiwan Fact Check Centre identified three common characteristics of this DMM: (1) unknown source, (2) use of simplified Chinese characters or Chinese idioms, and (3) to try and point out that the epidemic in Taiwan was out of control and accuse the central government of concealing the epidemic (Taiwan Fact Check Center, 2020).

Another example was when the WHO Director-General accused Taiwan of racial attacks against him, which resulted in a rapid influx of suspicious networked accounts purporting to be concerned Taiwanese internet citizens (aka netizens) apologising on behalf of Taiwan for their behaviour. Subsequent analysis by ASPI identified common patterns among the networked accounts; for example, following Chinese state associated accounts, recent creation dates, periods of inactivity and inconsistent activity, and retweets of known trolls or propaganda outlets (Thomas & Zhang, 2020).

The Poynter Institute's coronavirus database contained numerous China-attributed examples, such as the suggestion that an American soldier was intentionally spreading COVID-19 via saliva at a sports event in China (the actual photo was from Belgium and not of an American soldier); images showing Iranian senior officials visiting the country's vice president after she contracted COVID-19 (the actual image was from 2014 after being injured in a car accident); misleading images of a man who failed a body temperature test and attempted to force his way through a blockade (the actual image was of a police drill at a toll gate); photos showing a brothel in Europe where 86 people were being quarantined (the photo was of a nightclub in Marbella, with no-one quarantined); Italian people greeting and clapping for China after aid was delivered (the people clapping were supporting Italian healthcare staff during an organised flash mob—and was not related to China; however, the Chinese national anthem and a Chinese voiceover were added) (Wong et al., 2020); and a tweet by China's Global Times suggesting that Italy may have been the original source of the virus.

Finally, Filipinos were outraged and accused the Chinese of appropriating the pandemic for propaganda purposes when the Chinese embassy in Manila created a music video—*Iisang Dagat* (One Sea), which was meant to be dedicated to health workers from both countries in their collective struggle against the pandemic. Despite the Chinese trying to push a narrative about a new era of partnership and "one sea for everyone" the Filipinos reminded the Chinese that the disputed West Philippine Sea belongs to the Philippines after it won a 2016 case in the Permanent Court of Arbitration invalidating China's claims to almost the entire stretch of sea. Filipinos also reminded China that there certainly was not 'one sea' when the Chinese "point their guns and warships at our vessels and fishermen" (Reuters, DPA, 2020).

4.6 Terrorism and Violent Extremism

Violent extremist organisations have had at least a decade of experience using the internet and social media for creating and sharing DMM, and this also came to the fore during the pandemic (UN, 2020c). During the early phase of the pandemic ISIS blamed crusader nations in the West for the outbreak and urged followers not to travel to Europe to commit acts of terror, but instead suggested that followers in Iraq and Syria should attempt to free ISIS prisoners being held in camps (Hanna, 2020). The first narrative among the deluge of extremist online posts was related to God's retribution and punishment against corrupt and oppressive non-believers. One notable post called the virus the 'smallest soldier of God' (Meek, 2020). Another narrative, in order to explain why Muslims had died from COVID-19, was that they had been infected because they were not true Muslims (MEMRI, 2020). Clerics also cited Quranic verses that referenced "the importance of cleanliness, covering one's face when coughing and self-quarantining during a viral epidemic" (Hanna, 2020), suggesting that true Muslims were clean Muslims, therefore, immune from infection. In general, as with the activities of hostile state actors, the pandemic provided non-state actors and violent extremist groups with an opportunity to boost their central messaging (Kruglanski et al., 2020).

4.7 Readiness, Responsiveness and Operational Impact

It is clear that the pandemic had an effect on some aspects of military operational readiness and responsiveness; for example, the U.S. aircraft carrier USS Theodore Roosevelt, the Arleigh Burke-class guided-missile destroyer USS Kidd, and the French flag ship aircraft carrier Charles de Gaulle were all taken off operational missions and returned to port. The Theodore Roosevelt and Kidd sailed back to operational duties in June, whilst the Charles de Gaulle went back to sea in September 2020. At Camp Lemonnier in Djibouti a 30-day public health emergency was declared in late April 2020. Despite the narrative that it was a "precautionary measure", it attracted questions regarding operational readiness and responsiveness as the camp houses 3,000 U.S. military personnel, which is 50% of U.S. forces in the U.S. Africa Command. In Iraq, approximately 200 French forces suspended their training mission of Iraqi security forces in order to concentrate on supporting the national effort back home in France.

It is fair to assume that challenging events will happen again, whether the event is akin to a previous event or a new and unique event—and such events may occur sooner than anticipated; for example, in the UK, the gap between the Salisbury Novichok incident and coronavirus pandemic was only 24 months. To that effect, DMM is the "new normal" and will remain a fixture of future events, regardless if the context is health (e.g. pandemic), environmental and humanitarian (e.g. fires, drought, flooding, earthquake, volcano), political (elections) or economic (e.g. stock market crash).

4.8 Vaccine Hesitancy

The EUvsDisinformation website published an article explaining how pro-Kremlin media had consistently spread DMM regarding vaccine hesitancy/anti-vax/vaccine scare stories (EUvsDisinfo, 2020e). This is not a new tactic in the world of Russian disinformation and health events; for example, in 2019 the Russian DMM machine was accused of conducting a targeted campaign against Greece and its surge in measles cases as part of the global vaccine hesitancy problem (Kremidas-Courtney, 2019). Additionally, the Lugar Laboratory in Tbilisi, Georgia is accustomed to being a focus for pro-Kremlin DMM because it receives U.S. funding; therefore, it featured as part of a wider global set of DMM messages about the source of the virus emanating from western or western-backed laboratories around the globe (Table 4.3).

The context of health and medicine is also part of a longer running Russian influence campaign, especially in the Balkans, where there has been a growth in Russian-language alternative health magazines, particularly in Serbia, Slovenia, Croatia and North Macedonia. Not only do these magazines push alternative medicine 'cures' and anti-vax messages, but they do so through the prism of "Russian advice", with Russian symbology and flags prevalent throughout the magazine format. This is made all the more ironic as Russia has one of the lowest life expectancy rates in Europe (Global Voices, 2020).

The main focus for COVID-19 DMM during the spring and summer of 2020 was on the source of the virus, the causes of the spread, and cures and interventions (Table 4.3); however, as the media shifted its reporting toward a potential vaccine in the late summer and early autumn, so the DMM focus shifted more onto the race for a vaccine, as well as the expected vaccine hesitancy narratives and anti-vax movement. A key concern with this shift was the sheer potency of the anti-vax narrative; this is because it is such a dichotomous polarising issue (aka a 'wedge issue').

First, conspiracy theories began to be used. Jihadist extremists spread conspiracies that western governments would use a future vaccine to deliberately harm minority communities and control their populations. Ironically, both the far right and the jihadists used the 'sheeple' narrative (a contraction of 'sheep' and 'people'), which tried to reinforce the message about governments and elites (Commission for Countering Extremism, 2020).

By collecting a range of vaccine-related narratives during the pandemic, it can be seen how a single theme can be spun out into multiple alternative narratives (Fig. 4.1). This is an extension of the Bandwagon Effect (Table 4.2), but more dynamic and nuanced. It is a technique used by hostile states, socio-political movements (e.g. anti-vax), as well as regular trolls and sophists to perpetuate engagement on an issue. The perpetrator can jump from sub-narrative to sub-narrative, invent new narratives or recycle old ones. The aim is simply to perpetuate noise through engagement, i.e. if people are engaging and typing (aka keyboard warriors) then the issue is still 'live'.

In an effort to undermine and discredit the UK vaccine, domestic Russian media attempted to push a narrative of the 'monkey virus' (with associated memes), whereby the Russian state attempted to purport that the UK vaccine was based on an adenovirus from monkeys, whereas the Russian vaccine was based on human adenovirus. The narrative received hardly any traction in the British or European media, and only a limited amount of time in the domestic Russian media, or Russian-language media in regional East European countries, but it does highlight how the attack vector can rapidly and/or temporarily change within a topic—in this case the change in tactics, techniques and procedures (TTP) was to attack the quality of the UK vaccine versus the Russian vaccine, rather than the use of vaccines, per se. This simultaneously touches on aspects of economic warfare, as well as political warfare, via information

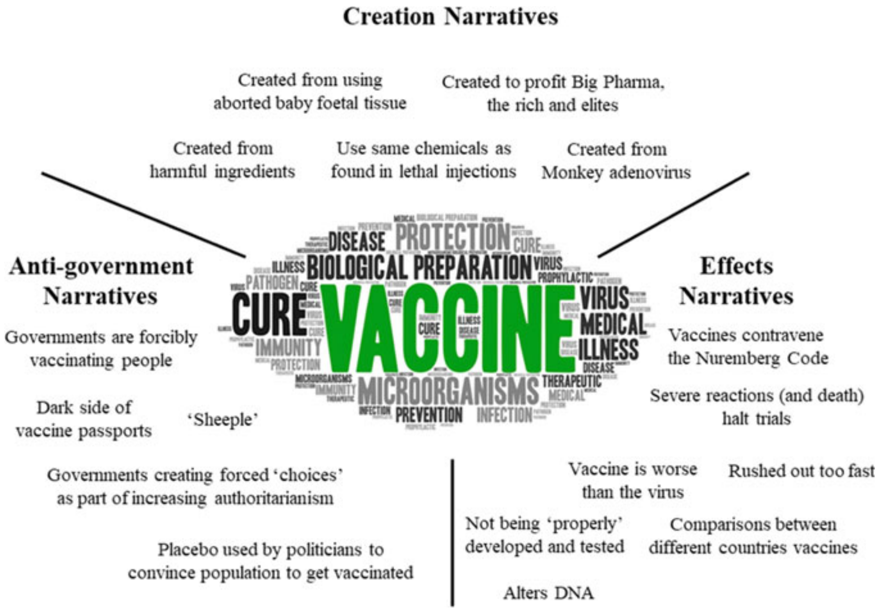

Creation Narratives
Created from using aborted baby foetal tissue
Created to profit Big Pharma, the rich and elites
Created from harmful ingredients
Use same chemicals as found in lethal injections
Created from Monkey adenovirus
Anti-government Narratives
Governments are forcibly vaccinating people
Dark side of vaccine passports
'Sheeple'
Governments creating forced 'choices' as part of increasing authoritarianism
Placebo used by politicians to convince population to get vaccinated
DISEASE
PROTECTION
CURE
ILLNESS
BIOLOGICAL PREPARATION
VIRUS
PROPHYLACTIC
PATHOGEN
CURE
VACCINE
VIRUS
MEDICAL
ILLNESS
IMMUNITY
PROTECTION
CURE
DISEASE
MICROORGANISMS
THERAPEUTIC
MEDICAL
PREVENTION
INFECTION
Effects Narratives
Vaccines contravene the Nuremberg Code
Severe reactions (and death) halt trials
Vaccine is worse than the virus
Rushed out too fast
Not being 'properly' developed and tested
Comparisons between different countries vaccines
Alters DNA

Fig. 4.1 Vaccine hesitancy narratives

means; and relates to three of the four instruments of power, i.e. the Diplomatic, Information, and Economic levers of the DIME structure (with M for Military).

Research by the London School of Hygiene and Tropical Medicine suggested that susceptibility to DMM reduces a person's intention to be vaccinated, in both representative UK and U.S samples (Loomba et al., 2021). These sorts of findings can be used to provide explanatory power for reductions in vaccine compliance rates, as well as for wider issues such as staying at home, obeying physical distancing rules, wearing masks in public, and the growth of the anti-lockdown movement. It is interesting to note that some well-known individuals for one issue also became known for COVID-related issues; for example, in the UK, the well-known climate change denier Piers Corbyn also became known for anti-vax and anti-lockdown. Additionally, in the UK in January 2021, anti-lockdown activists were accused of stage-managing an incident by enticing the arrest of a woman by Police for breaking lockdown rules. The incident was filmed by anti-lockdown activists and disseminated across social media (Powell, 2021).

4.9 UK Capability and Intervention

4.9.1 Building a Capability

Since 2017 the UK's Government and Parliament have produced several key publications (Table 4.5). The Department for Digital, Culture, Media and Sport (DCMS) is the lead department for the broader issue of online harms, which includes DMM.

Table 4.5 UK Government core disinformation documents and guidance

Year	Publication
2017	Internet Safety Strategy Green Paper (HM Government, 2017)
2018	House of Commons Interim Fake News Report (House of Commons, 2018)
2019	House of Commons Final Report on Disinformation (House of Commons, 2019)
2019	Online Harms White Paper (HM Government, 2019a)
2019	RESIST counter-disinformation toolkit (GCS, 2019)
2019	Privacy Notice—analysis of social media data to tackle disinformation (EU and US partners) (HM Government, 2019b)
2020	Interim code of practice on Terrorist Content and Activity Online (Home Office, 2020)

DCMS also leads on the policy response to countering disinformation and coordinates the various efforts across the UK government. Prior to the pandemic, certain government departments, including the Home Office, Cabinet Office and Foreign, Commonwealth and Development Office (FCDO), possessed their own counter-disinformation teams. Their roles are to address day-to-day issues of DMM relevant to their department, but they may also be operationalised through the DCMS-led Counter-Disinformation Unit for high-profile national events, such as the 2019 EU Parliament Elections and the 2019 UK General Election.

The UK has also been very active in creating and supporting international partnerships; most notably with the G7 Rapid Response Mechanism, including engagement with the European Union (EU) via the EU Commission and the EEAS; and also with the U.S via the State Department. In 2018 the government also launched an online campaign called Don't Feed the Beast, which aimed to educate and empower those who see, inadvertently share and are affected by false and misleading information. This included the SHARE checklist (Cabinet Office, 2020; HM Government, 2020a).

- **S**ource: make sure information comes from a trusted source;
- **H**eadline: always read beyond the headline;
- **A**nalyse: check the facts;
- **R**etouched: does the image or video look as though it has been doctored?
- **E**rror: look out for bad grammar and spelling.

Government efforts to tackle DMM have also been supported by other UK initiatives, such as those initiated by civil society; for example, Full Fact is an independent fact-checking charity that was established in 2010, and is part of the Poynter Institute's IFCN; and the British Broadcasting Corporation (BBC) created its Reality Check team in 2017.

In 2018 researchers at the University of Cambridge created the Bad News Game, which is a fake news intervention aimed at building psychological resistance against online misinformation. The game is described as a "theory-driven social impact game" and in collaboration with the FCDO has been translated into15 languages, as well as being supported by several peer-reviewed publications (Basol et al., 2020; Maertens et al., 2021; Roozenbeek & van der Linden, 2018; Roozenbeek et al., 2020; van der Linden & Roozenbeek, 2020). Results from their research show that a single play of the game can inoculate players for up to three months, and that susceptibility to DMM was reduced by 21%.

In 2019 the University of Oxford established the International Multimodal Communications Centre. The centre is a pan-Oxford partnership, including the Oxford Internet Institute, which conducts ground-breaking research into multimodal communications, including machine learning and automation, which includes research into DMM. In 2020, the National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online (REPHRAIN) was launched. The centre is a collaborative effort between the universities of Bristol (lead university), University College London, King's College London, Bath and Edinburgh. The Centre has three missions: Mission 1: Automated methods into novel text mining and image/video analysis; Mission 2: Maximising benefits from a sharing-driven digital economy; Mission 3: The balance between individual agency versus social good.

4.9.2 Intervention During the Pandemic

The UK's attention regarding DMM for COVID-19 began in January 2020. This attention became more focused as the nature of the pandemic became clearer. This resulted in the cross-Whitehall (XWH) Counter Disinformation Unit (CDU) being stood up on 5 March 2020. The CDU brought together various cross-Government monitoring and analysis capabilities, whose primary function was to provide a comprehensive picture of the extent, scope and reach of DMM linked to COVID-19, and to work with partners to remove and mitigate it. Also in March, the UK Parliament re-established the Sub-Committee on Online Harms and Disinformation, as they considered the growing issue of DMM related to COVID-19 (UK Parliament, 2020). On 30 March a press release was issued by the UK government, which explained the XWH coordination across the government department teams (HM Government, 2020b). The government's response also included the British Army's 77th Brigade, which was tasked to support wider government under Operation RESCRIPT, which was the British military's efforts to support government in its response to the pandemic (Young, 2020), as well as the British military's support to NATO (HM Government, 2020c).

The government's counter-disinformation efforts produced Situation Reports (SITREPS) and updates through various existing reporting mechanisms. These started as weekly reports and products at the start of 2020, but it rapidly became apparent that daily SITREPs were required as the volume of DMM exponentially increased, both nationally and globally. The reporting mechanisms were consolidated in order to reduce duplication so that a mix of daily SITREPS and consolidated weekly updates were operating in concert. When the tempo began to slow into the autumn the reporting reverted to weekly products, but this still continued in 2021, particularly with anti-vax narratives and numerous country-specific variants that circulated.

The UK government also worked closely with social media platforms throughout the pandemic to help them identify and take action to remove incorrect claims about the virus, in line with their revised terms and conditions, and to promote authoritative sources of information. The DCMS Secretary of State met with the major platforms on a number of occasions to bring together social media companies, civil society groups and health experts. Building on these discussions, the government convened regular Counter-Disinformation Policy Forums, which brought together key actors in industry, civil society and academia to improve responses to DMM and prepare for future threats. The Forum informed their understanding of challenges to the DMM ecosystem; as well as helped improve responses to the evolving threat posed by DMM.

The Office for Communication (Ofcom), who are the regulatory and competition authority for the broadcasting, telecommunications and postal industries of the UK, published a range of information on how people were getting news and information about COVID-19 (Ofcom, 2020a), as well as a survey of views about misinformation (Ofcom, 2020b). Government communications professionals already possessed the RESIST toolkit and the government relaunched the Don't Feed the Beast campaign to support early COVID-19 mitigations.

The University of Cambridge team, who had created the Bad News Game, worked in collaboration with the UK Cabinet Office to create a new game for the pandemic. The 'Go Viral' game looked at pre-bunking COVID-related DMM as a form of information inoculation. The game has already been translated into German and French, with more translations and cross-cultural research comparisons due to take place in 2021.

Overall, it can be seen that the UK government has invested heavily in countering disinformation since 2016, in terms of intellectual capacity, policy, and response cells. The journey has been one of steady improvement, using an 'action research' model of iterative improvement, whereby it has tested and adjusted its capabilities and processes on various elections and events. This was fortuitous when the pandemic hit the UK in early 2020, whereby experience and lessons learned could be focused onto the pandemic and the associated, and inevitable, DMM.

4.10 Disinformation as a System

To assist governments and international organisations in tackling DMM, it is necessary to start viewing DMM as a planned, deliberate and targeted system by hostile states and threat actors, rather than the pre-2016 assumptions of a random set of unconnected and uncoordinated trolls using a scatter gun approach. This is because adversaries have adapted their DMM TTPs (aka tradecraft), and they are now acutely aware that over-simplistic fake facts can be quickly and easily identified, and robustly debunked by the IFCN; therefore, adversaries have become smarter and more nuanced in their TTPs. Thus, by using a 'systems thinking' approach, it is possible to gain an improved understanding of how DMM is evolving.

A systemic approach to DMM was seen during the first wave of the pandemic whereby a "trilateral convergence of disinformation narratives [was] being promoted by China, Iran, and Russia" about the pandemic (Jozwiak, 2020b). This not only involved hostile states creating their own specific disinformation messages and memes, but also the re-use and amplification from other hostile states for mutual interests, e.g. sanctions relief.

An evolving technique within this system is to re-share and amplify the narratives of other sources rather than be the creator of them—thereby pleading plausible deniability about attribution and origin of source. This falls short of *mis*information because the passing on of information is deliberate rather than inadvertent, and because it is belligerent with an intent to harm (or influence) it fits the description of malinformation. An example of this during the pandemic was when a left-wing Italian paper published an article criticising NATO's Defender Europe 20 military exercise, which was being conducted at the same time that the first wave was escalating in Europe (Dinucci, 2020). This narrative was picked up by Russian media, translated into Russian, and spread across Russia-speaking mainstream media and wider social media (Andriukaitis & Pellegatta, 2020).

An old strategy that has re-emerged is the 'agents of influence' (Schultz & Godson, 1984). This strategy has been updated for the twenty-first century and includes the involvement and coordination of a range of individuals and organisations, e.g. digital, media, political, civil society, think tanks and academia (Fig. 4.2). The nexus coordinates and collaborates amongst itself to promote their own narratives and/or support individuals linked to their political and economic interests, whist simultaneously undermining and discrediting opponents (Shekhovtsov, 2020). This was seen during the pandemic whereby mutual following, re-publishing and re-tweeting COVID-related DMM not only helped to propel the DMM wider, but also helped to perpetuate it across time. Furthermore, individuals or organisations within the nexus appear to provide a veneer of respectability, whilst obscuring their links to state-funded entities (GEC, 2020a). Increasingly, the nexus engages in global partnerships, where there is graded coordination that ranges between centralised, quasi-independent, independent; as well as blended authenticity to help convince the audience (Decker, 2019).

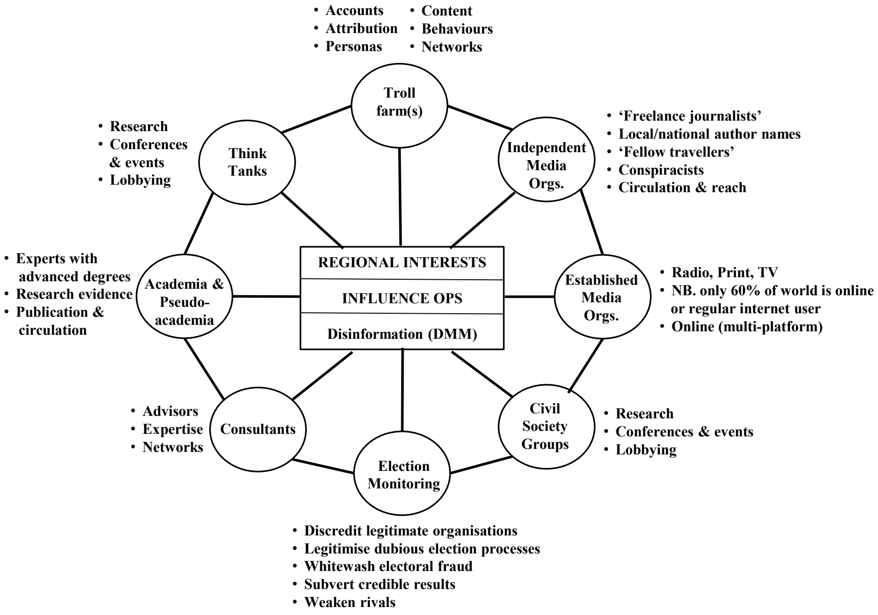


Fig. 4.2 Network of agents of influence

4.10.1 Funding Disinformation

The Global Disinformation Index (GDI) published several reports and a weekly tracker with insights into how the pandemic was being exploited by generating illicit revenue through the manipulation of advertising technology (ad tech). In fact, analysis by GDI claims that within Europe alone "disinformation news sites take in more than US$76 million each year in revenues from online ads" (GDI, 2020). Furthermore, analysis by EUvsDisinformation found that certain pro-Russian media outlets (who also utilise ad tech) used COVID-19 disinformation and conspiracy theories as a way of increasing web traffic by as much as 258% (EUvsDisinfo, 2020f, 2020g). Analysis on COVID-19 ad tech by GDI found that:

> In a sample survey of nearly 50 sites carrying coronavirus conspiracies, Google provided ad services to 86% of these sites and Amazon 20%, i.e. ad tech players are serving up ads and providing ad revenue streams to known disinformation sites peddling coronavirus conspiracies. While ad tech players are making great efforts to tackle coronavirus content on their platforms, they have more work to do in cutting the ad services they supply to the very content they are trying to remove. By providing ad services to these sites that promote coronavirus conspiracy theories, ad tech companies are inadvertently funding and fuelling the problem. (GEC, 2020a)

Interestingly, in 2018 the former Director of the UK's Government Communications Headquarters (GCHQ) gave a keynote speech at the Infosecurity Europe conference to highlight how cybercrime groups and hostile state actors have blurred together and are jointly using creativity, innovation and agility to weaponise the World Wide Web for criminal gain and nation state objectives.

> In some cases, you can see these groups sitting in the same room, and in some cases, you can see where people have been conducting state activity during the day, and then doing crime activity at night…It's an interesting mixture of profit and political intent. (Schwartz, 2020)

Figure 4.3 visualises this interplay in terms of disinformation and ad tech whereby: (1) disinformation content (aimed at COVID-19) is created. Then (2) fake internet traffic is generated, thereby attracting (3) real internet traffic; which in turn (4) attracts ad tech, thereby (5) creating revenue for both (6) the crime gang and/or the state/non-state actor.

4.11 Future Threats from Disinformation

It is becoming increasingly apparent that the topic of disinformation and misinformation is not going to retreat into the background, as it did in the West after the Cold War. It has also been suggested that the world is moving towards its first 'information civilization' (Zuboff, 2015), therefore, it is fair to hypothesise that DMM will be a part of the global digital-online landscape for the foreseeable future. Two key factors will shape an information civilization: (1) the grand strategic human picture, and (2) our future digital lives.

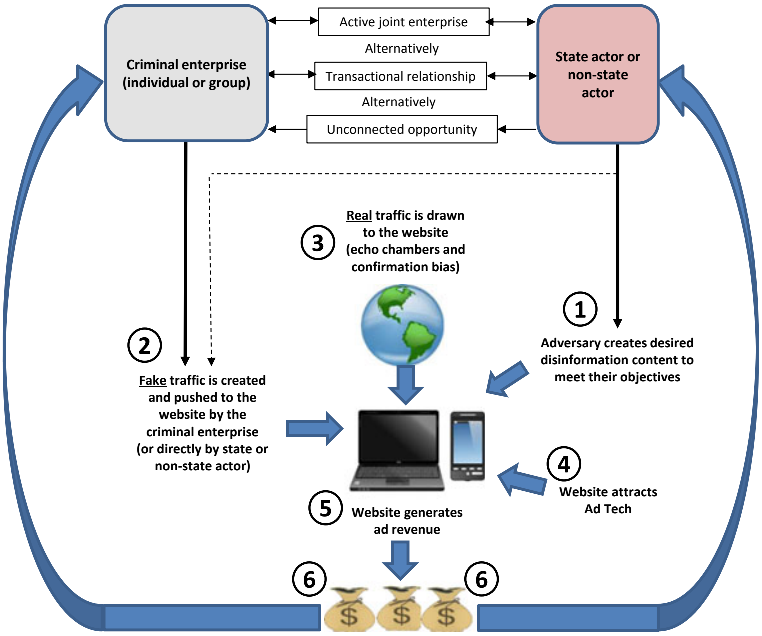

Active joint enterprise
Alternatively
Transactional relationship
Alternatively
Unconnected opportunity
Criminal enterprise (individual or group)
State actor or non-state actor
3 Real traffic is drawn to the website (echo chambers and confirmation bias)
1 Adversary creates desired disinformation content to meet their objectives
2 Fake traffic is created and pushed to the website by the criminal enterprise (or directly by state or non-state actor)
4 Website attracts Ad Tech
5 Website generates ad revenue
6
6
$
$
$

Fig. 4.3 The disinformation ad tech complex

4.11.1 The Grand Strategic Human Picture

Human movement on planet earth is changing; whereby a range of human security issues are having powerful effects; for example, water security, food security, energy security, health security, as well as climate change and environmental issues. The world’s population at the time of writing this chapter was 7.8 billion. The population is expected to begin to plateau at 11 billion, by the end of the century in 2100. The majority of this population growth is expected in Asia and Africa. On top of population growth there will also be significant human movement and migration. People are increasingly moving to urban conurbations; whereby it is assessed that 68% of the world’s population will be living in the urban environment by 2050, creating vast urban sprawls or ‘megacities’. It is estimated that by 2050 there will be somewhere in the region of 50 megacities, mostly in Asia and Africa; with the individual population of each of the top ten megacities ranging between 35 and 50 million inhabitants—with the vast majority of these people wanting to be, or needing to be, digitally connected (Verrall, 2019).

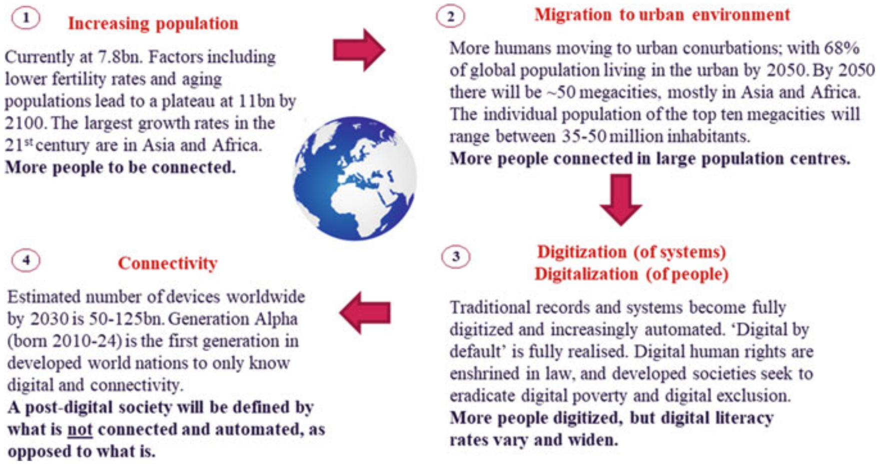
4.11.2 Digital Lives

Future digital lives will be driven by changing societies and societal expectations, as well as digital economies and how governments provide digital governance for their populations. The current Generation Alpha is the first generation in history where the majority live their lives (and expectations) as digital by default. As older generations die off and developing nations become increasingly digital, future generations (Generation Beta and beyond) will form part of super- (or even hyper-) connected nations and societies, where smart cities and the people in those cities are connected by the internet of things at all levels, from the individual to the national. The continued rise of individualism and fragmentation of personal identity will increase the demand for individual agency in the digital domain, whereby individuals will become increasingly empowered as they seek to fulfil their desire for more, and better, engagement and experience.

Additionally, digital human rights will become an important aspect of people's daily lives. Digital human rights centres around three main issues: (1) the right to be connected. This is already occurring in the younger generations, whereby they see the ability to connect and have access to the internet as an inalienable human right. Research has also identified this concept within younger military samples (Adey et al., 2016), (2) the right to erasure (aka the right to be forgotten). This links to issues of individual agency and empowerment, whereby the individual will have more of a say about their data (which they own), rather than data about them, that is currently stored by the internet companies, (3) the aspiration of governments and civil society groups to eradicate digital exclusion, thereby helping to reduce digital poverty. It should be remembered that in 2020 only 60% of the world's population were active users of the internet (Statista, 2020), but this percentage will only increase over time. Ultimately, a digital life will be by default and will be noticed by what is digitally absent, rather than what is digitally present or possible—this is the emergence of the post-digital society (Negroponte, 1998). Figure 4.4 illustrates this human and technology journey (Verrall, 2019, 2021).

4.11.3 Future Technologies and Future Threats

The Netflix programme, the *Social Dilemma*, provides a good example of what happens when computer scientists and tech entrepreneurs focus on the upside of new and innovative technology, and do not think sufficiently about the downside and the risk. The same can be applied to DMM, whereby emerging technologies, as well as developing adaptions and innovations to existing technologies, will potentially provide numerous attack vectors for future DMM by a range of threat actors, such as hostile state actors, hostile non-state actors (violent extremist organisations and terrorist groups), proxy forces, organised crime groups and political activists (Verrall, 2021). Examples of developing and emerging technologies include:



Currently at 7.8bn. Factors including lower fertility rates and aging populations lead to a plateau at 11bn by 2100. The largest growth rates in the 21st century are in Asia and Africa.
More people to be connected.

More humans moving to urban conurbations; with 68% of global population living in the urban by 2050. By 2050 there will be ~50 megacities, mostly in Asia and Africa. The individual population of the top ten megacities will range between 35-50 million inhabitants.
More people connected in large population centres.

Traditional records and systems become fully digitized and increasingly automated. 'Digital by default' is fully realised. Digital human rights are enshrined in law, and developed societies seek to eradicate digital poverty and digital exclusion.
More people digitized, but digital literacy rates vary and widen.

Estimated number of devices worldwide by 2030 is 50-125bn. Generation Alpha (born 2010-24) is the first generation in developed world nations to only know digital and connectivity.
A post-digital society will be defined by what is not connected and automated, as opposed to what is.

Fig. 4.4 The changing nature of humans and technology

- **The internet of things**. This describes a world in which everyday networked devices are defined by connection, collection, computation and creation.
- **Decentralised internet**. A move away from internet domination by a small number of hosting companies towards a more decentralised internet that is people-powered and allows for a more democratic and fair model of opportunity and growth.
- **Privacy enhancing technologies**. Methods of hard and soft technology to protect privacy and personally identifiable information.
- **Artificial intelligence and machine learning**. Algorithms (code) that can process and act independently of human involvement.
- **Automation and autonomy**. A system that has been instructed to perform specific tasks within specified parameters, independent of human involvement (i.e. automation); as well as when that system subsequently determines its own course of action to achieve defined outcomes (i.e. autonomy).
- **Affective computing**. Use of wearable devices to enhance the direct communication between technologies and the brain (brain-computer interaction - BCI) and how BCI can be exploited to enhance communication between human brains without periphery nervous systems, i.e. brain-to-brain communication (B2B).
- **Digital immersion**. Developments in areas such as virtual reality, augmented reality and tactile internet that enable the real-time transmission of haptic information.

There are numerous ways in which threat actors can exploit these technologies by displaying agility and innovation in order to identify opportunities, fissures and loopholes to create and/or disseminate DMM (Table 4.6). Their motivational drive is to develop the ability to:

Table 4.6 How technology can be exploited for DMM threat

Technology trend	Threat opportunity
Internet of things	Hacking into everyday devices, e.g. swatting behaviour. IoT networks can be used as a form of 'super-spreader' of DMM, allowing sophists to deliver DMM anytime, anywhere, on any device, on any network, at any individual, group or organisation, in any real world context or event
Decentralised internet	A flatter, distributed, decentralised internet allows DMM to be created and proliferate within the inefficient margins of traditional command, control and communication models
Privacy enhancing technologies	Allows sophists to hide their identity and location; thereby stymieing governments' ability to establish accurate attribution, which may impact ability to target for mitigation and disruption purposes
Artificial intelligence, machine learning, automation and autonomy	The ability to operate 24/7, at scale and speed, by enabling DMM software to FIND (who to send it to), COLLECT (what to collect), ANALYSE (what to analyse), CREATE (what to create), TARGET (who to target), DISSEMINATE (how to best share), NETWORK (how to identify the best networks to share)
Affective computing and digital immersion	Exploiting BCI and B2B to accelerate the speed and dissemination of DMM to individuals and groups. Can be exploited to collect and understand as well as manipulate and influence
Attention economics	Human attention increasingly becomes a valuable resource in the digital age because attention is the limiting factor, not information
Disinformation as a service	The supply and demand of disinformation skills and capabilities to sophist customers
Intelligence as a service	Threat actors have access to, and demonstrate, military-grade Intelligence tradecraft and capabilities in order to Direct, Collect, Process and Disseminate (aka the intelligence cycle)
Future digital-online media landscape	The increased fragmentation of traditional media into multi-partisan options for news and media
Digital and media lawfare	Utilising and exploiting legal and regulatory loopholes to gain an advantage. International and national laws can be exploited to stop, delay or circumvent
Adaptive tradecraft	Adversaries, threat actors and sophists continue to demonstrate agility and innovation to adapt DMM tradecraft (Tables 4.1 and 4.2)

- Target any device on any network path;
- Target anybody, place, business or organisation;
- Target anywhere, anytime;
- Apply this to any context or event (e.g. elections and political events, health events [pandemic], environmental event, stock market crash, military intervention, etc.).

Potential societal responses to such threats will be numerous and varied. Examples of some responses include:

- The long-term erosion of confidence and trust in institutions, governments and experts;
- A proliferation of polarised echo chambers of individual certainty, whereby individuals who are looking for identity and belonging are pushed to the radical and extreme peripheries of certain wedge issues;
- The emergence of socio-digital dissonance, which could be described as a state of psychosocial conflict within individuals between two competing tensions: (1) a preference for ease of connectivity and instant access to the things that make life quicker, simpler, better (e.g. entertainment, browsing and shopping, banking, finding and learning, communicating, online working, social networking, etc.), versus (2) the need to be responsible for the relatively mundane and time-consuming threats and risks regarding online security;
- Cognitive abilities will be challenged in multiple ways, e.g. the capacity for analytical thinking, memory, focus, creativity, reflection and mental resilience (incl. mental health and well-being);
- The risk-centric consumer, whereby internet companies will shift the risk of data and information vulnerability away from themselves and onto the consumer as a way of being seen to fulfil the consumer's desire for agency, empowerment, engagement and experience.

Therefore, it can be seen how a future information civilization, in a post-digital society, may also end up in a situation of information disorder, if threats from developing and emerging technologies are exploited by threat actors for the purposes of DMM (Fig. 4.5). Some would argue that we are already in a state of information disorder (Wardle, 2019), whereby people are already subjected, on a daily basis, to disinformation, misinformation and malinformation, as well as information overload and competing interpretations of information that skew, obfuscate and influence people's level of understanding and knowledge.

4.12 Impact on Defence and Security

Earlier in the chapter it was mentioned how COVID-19 and DMM could impact operational readiness and reaction. At the heart of this is the growing appreciation of how DMM contributes to contemporary concepts of warfare such as hybrid warfare and grey zone activities. DMM predominantly operates below the traditional threshold of

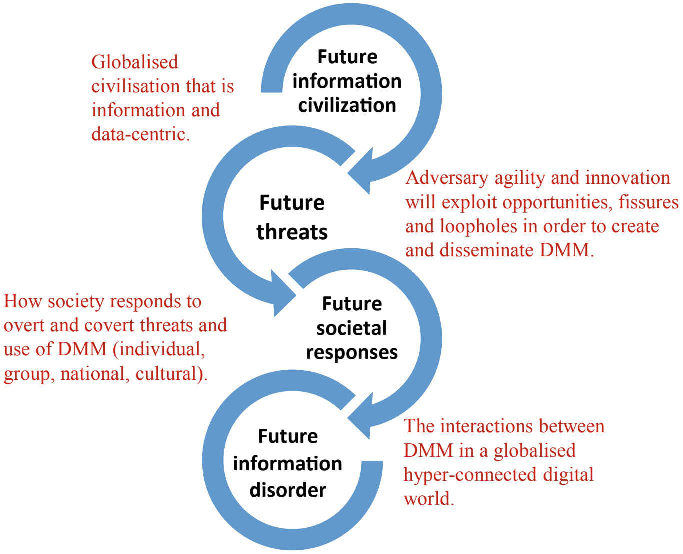

Future information civilization
Globalised civilisation that is information and data-centric.
Future threats
Adversary agility and innovation will exploit opportunities, fissures and loopholes in order to create and disseminate DMM.
Future societal responses
How society responds to overt and covert threats and use of DMM (individual, group, national, cultural).
Future information disorder
The interactions between DMM in a globalised hyper-connected digital world.

Fig. 4.5 From information civilization to information disorder

warfighting (aka the sub-threshold), where threat actors can enjoy a greater [digital] freedom of manoeuvre and wider attack space via the internet. This enables threat actors to be innovative, agile and diffuse. Two UK military doctrine publications understand this changing threat—the Integrated Operating Concept (UK Ministry of Defence [MOD], 2020) and Information Advantage (UK MOD, 2018). In the Integrated Operating Concept there is recognition that military forces will increasingly take part in activities that are short of warfighting, and will operate at the sub-threshold level in order to protect the homeland, as well as engage and constrain adversaries and threat actors across the globe. Information Advantage doctrine articulates how information and Information Systems will act as enablers and affecters in the defence and security domain. It honestly and openly recognises that the military does not currently hold the initiative when it comes to certain issues, including DMM. This was demonstrated during the early phase of the pandemic when some military forces were sometimes slow to understand and react to the situation; for example, aggressive counter-narratives about the Defender Europe 20 exercise, NATO's slow response to assist allies, and reductions in operational manpower due to outbreaks or re-prioritization. The pandemic has only solidified to the military that in an era of persistent competition, where belligerent activities often occur below the traditional threshold of war, that all, and any, opportunity will be exploited by adversaries in the digital-online domain in order to undermine trust and discredit reputation and credibility (HM Government, 2021). The MOD recognises this and is in the process of producing a Joint Concept Note on Sub-Threshold Advantage (UK MOD, 2022), supported by the Information Advantage doctrine, which posits that the military does have a role in supporting the wider government apparatus in a whole-of-government approach, by employing capabilities that help to sense and shape the environment.

4.13 Discussion and Conclusion

The news is not all bad. It is worth mentioning that governments can anticipate certain threats and develop mitigations; this includes utilising developing and emerging technologies for the benefit of counter-DMM capabilities. Other activities include improving the standards, codes and regulations that create the loopholes for threat actors to enjoy the freedom of manoeuvre in a digital environment. Also, devices that will be part of the Internet of Things must be 'secure by design'. Some of this will have no doubt been in place for the pandemic (e.g. improved social media terms of service, improved data analytics) but governments tend to take time to bring their counter-capabilities up-to-speed, before eventually over-matching smaller or weaker threat actors. However, in a fast-paced online environment, DMM is as much about speed as it is about reach and engagement. It is also worth remembering that the problem of standing-up to hostile state actors is not necessarily the issue of capacity, speed and scale (as it is with non-state threat actors), the problem is more closely linked to will, i.e. the political desire to pre-empt, to monitor and to act when needed. Professional militaries refer to this as the moral component of fighting power (UK MOD, 2014).

The DMM curve peaked in April 2020, however, its reduction continued on a slow, long tail for many months thereafter (and into 2021) as the virus shifted its continental epicentres, as well as the context and narratives of COVID-related DMM also shifting, i.e. cause, spread, cure, vaccines, freedoms, and economics. Some commentators remarked rather early (June 2020) that "the subject is no longer as conducive for disinformation" (Schwartz, 2020); however, this proved premature because of the shifting narratives towards vaccines, lockdowns and economics; specifically, the anti-vax and anti-lockdown movements. The vast amount of DMM that emerged demonstrated the diffuse nature of the problem. Lots of people, in many countries, had their say and shared their opinions—from troll, to gifted amateur, to expert, to politician. The DMM ecosystem is no longer about obvious fake facts per se, but now involves strategies; agents of influence networks; ever-changing TTPs; the exploitation of ad tech to generate revenue; the interplay between hostile state actors, threat actors…and good old fashioned opportunism by trolls and conspiracy theorists.

Facebook suggested that COVID-19 provided an opportunity for hostile states to piggyback onto existing pre-pandemic narratives (e.g. anti-West, anti-NATO, anti-American, anti-Elites) rather than COVID-19 being a bespoke, planned line of operation (Facebook, 2020). Eventually, hostile states reverted back to pre-pandemic DMM themes and narratives, particularly from Russia, e.g. the West is no longer interested in Ukraine, and the collapse of Western values, the increasing encroachment of NATO; however, new phases of the pandemic provided new opportunities to remain active in DMM, as seen by the creation of vaccines in late 2020, new variants of COVID-19 and new lockdowns in early 2021, and the roll out of vaccinations throughout 2021.

The UK's counter-disinformation response to COVID-19 was based on a capability that has steadily improved since 2017. Ironically, the global exposé of Russian interference in the 2016 U.S Presidential Election provided the impetus to address DMM by international governments, which meant that many nations were better prepared for the infodemic that accompanied the pandemic than they might otherwise have been. One can only imagine the global chaos and challenge of having to deal with COVID-19 DMM if nations had to deal with it from a standing start.

It is also interesting to note the expanding lexicon of misinformation and disinformation language. The dominant terms for decades were both misinformation and disinformation. This has been supplemented with malinformation and information disorder. The WHO have now introduced infodemic and also infodemiology, which is similar to epidemiology in that it seeks to study, understand and analyse the distribution, patterns and determinants of health-related information (Eysenbach, 2002), whereas traditional epidemiology seeks do achieve the same, but for medical and health conditions. As information and DMM expands in our daily lives we should expect to see the lexicon of DMM and information disorder to also change.

The future of DMM must be considered. We are moving rapidly toward an information civilization, which is prone to information disorder. Notwithstanding the efficient and pro-social benefits that emerging technologies are posited to provide, they will also be exploited by hostile states and threat actors—as it was ever thus. The worst case scenario for this outlook is where threat actors can operationalise DMM by targeting any device against anybody, anywhere, anytime—automatically, autonomously and ubiquitously. Interestingly, when platform content moderators were sent home due to governments' physical distancing interventions in the first lockdown in early 2020, there was an immediate reduction in the performance efficiency of machine learning algorithms that did not have consistent human oversight (Sharwood, 2020). This not only works in the favour of the sophists, but also demonstrates that not only are algorithms and machine learning still highly fallible, but that human–machine teaming is likely to (and should) remain with us for the longer-term, i.e. leaping to full automation and autonomy may help to solve some problems in society, but it will create others, especially where DMM is concerned. This is the lesson that computer scientists, technologists and tech entrepreneurs failed to fully consider and appreciate the first time around.

In conclusion, DMM never really went away after the Cold War, but it is definitely here to stay. The internet, online news media and social media offers opportunities for the creation, reach, speed and engagement with DMM that far exceeds those of the fifteenth century printing press, wireless, radio and television. COVID-19 provided a new attack vector for the peddlers of sophistry, and we should expect DMM to play a part in any future event (foreseen or unforeseen), be that political, economic, environmental, medical or technological. If the present society wishes to negate the looming information disorder within future society, then all it takes for the sophists and threat actors to succeed with DMM is for decent people, ethical industries and responsible governments to not take the subject matter seriously and/or mitigate it. If we have learned anything in the first quarter of conflict and competition this century, it is that menace loves an opportunity, and there appears to be many opportunities amongst the emotional, political, social and informational vacuums that affect the lives of contemporary human beings.

References

Adey, P., Denney, D., Jensen, R., & Pinkerton, A. (2016). Blurred lines: Intimacy, mobility, and the social military. *Critical Military Studies, 2*(1–2), 7–24. https://doi.org/10.1080/23337486.2016.1148281

Agenda. (2020a, February 28). *Russia to send mobile brigades to occupied Abkhazia to prevent coronavirus*. https://agenda.ge/en/news/2020/617

Agenda. (2020b, April 18). *Int'l community condemns Russian 'borderization' in Georgia*. https://agenda.ge/en/news/2020/1205

Aleksejeva, N. (2020a, April 29). *Fact-checker's identity stolen to spread disinfo about NATO and COVID-19*. Medium.com. https://medium.com/dfrlab/fact-checkers-identity-stolen-to-spread-disinfo-about-nato-and-COVID-19-111a2eef70a0

Aleksejeva, M. (2020b, April 17). *COVID-19 conspiracists conspire against fact-checkers*. Medium.com. https://medium.com/dfrlab/COVID-19-conspiracists-conspire-against-fact-checkers-c1909130a008

Andriukaitis, L. (2020, March 28). *Pro-Kremlin media spins story of U.S. military transporting COVID-19 test swabs from Italy*. Medium.com. https://medium.com/dfrlab/pro-kremlin-media-spins-story-of-u-s-military-transporting-COVID-19-test-swabs-from-italy-548b98c0435d

Andriukaitis, L., & Pellegatta, A. (2020, March 24). *Italian anti-NATO coronavirus narrative recycled in Russia*. https://medium.com/dfrlab/italian-anti-nato-coronavirus-narrative-recycled-in-russia-46f14537c25a

ASPI. (2020, April 23). *COVID-19 disinformation and social media manipulation trends 8 April–15 April*.

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition, 3*(1), 1–9.

Bittman, L. (1985). *The KGB and Soviet disinformation: An insider's view*. Pergamon Brassey's.

Bjola, C., & Pamment, J. (2016). Digital containment: Revisiting containment strategy in the digital age. *Global Affairs, 2*(2), 131–142. https://doi.org/10.1080/23340460.2016.1182244

Boghardt, T. (2009). Operation INFEKTION—Soviet bloc intelligence and its AIDS disinformation campaign. *Studies in Intelligence, 53*(4), 1–24.

Braw, E. (2020, March 30). Beware of bad Samaritans. *Foreign Policy*. https://foreignpolicy.com/2020/03/30/russia-china-coronavirus-geopolitics/

Cabinet Office. (2020, January 2). *Freedom of information request: Don't feed the beast FOI329095*. Cabinet Office.

Center for European Policy Analysis. (2020). *Disinformation techniques*. https://www.cepa.org/disinfo-techniques

CIA. (1986). *The Soviet foreign propaganda apparatus*. CIA.

Commission for Countering Extremism. (2020). *COVID-19: How hateful extremists are exploiting the pandemic*. CCE.

Decker, B. (2019). *Adversarial narratives: A new model for disinformation*. Global Disinformation Index.

Dinucci, M. (2020, March 3). 30 thousand soldiers from the USA in Europe without a mask. *Il Manifesto*. https://ilmanifesto.it/30mila-soldati-dagli-usa-in-europa-senza-mascherina/

EUvsDisinfo. (2019, July 18). *Tracing five years of pro-Kremlin disinformation about MH17*. EUvsDisinfo. https://euvsdisinfo.eu/tracing-five-years-of-pro-kremlin-disinformation-about-mh17/

EUvsDisinfo. (2020a, March 19). *EEAS SPECIAL REPORT: Disinformation on the coronavirus—Short assessment of the information environment*. https://euvsdisinfo.eu/eeas-special-report-disinformation-on-the-coronavirus-short-assessment-of-the-information-environment/

EUvsDisinfo. (2020b, April 30). *Flattening the curve*. EUvsDisinfo. https://euvsdisinfo.eu/flattening-the-curve/

EUvsDisinfo. (2020c, May 4). *COVID-19: Pro-Kremlin media defend Beijing*. EUvsDisinfo. https://euvsdisinfo.eu/COVID-19-pro-kremlin-media-defend-beijing/

EUvsDisinfo. (2020d, April 8). *Coronavirus disinformation: Moscow overplays its hand*. EUvsDisinfo. https://euvsdisinfo.eu/coronavirus-disinformation-moscow-overplays-its-hand/

EUvsDisinfo. (2020e, April 16). *Vaccine hesitancy and pro-Kremlin opportunism*. EUvsDisinfo. https://euvsdisinfo.eu/vaccine-hesitancy-and-pro-kremlin-opportunism/

EUvsDisinfo. (2020f, May 4). *New stories, old plots*. EUvsDisinfo. https://euvsdisinfo.eu/new-stories-old-plots/

EUvsDisinfo. (2020g, May 5). *Figure of the week: 117 million*. EUvsDisinfo. https://euvsdisinfo.eu/figure-of-the-week-117-millions/

Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis)information. *American Journal of Medicine, 113*(9), 763–765. https://doi.org/10.1016/S0002-9343(02)01473-0

Facebook. (2020, May 5). *Coordinated inauthentic behavior report*. Facebook. https://about.fb.com/news/2020/05/april-cib-report/

GEC. (2020a). *Special report: Pillars of Russia's disinformation and propaganda ecosystem*. GEC.

GEC. (2020b, January 29). *Russian disinformation apparatus taking advantage of coronavirus concerns*. Global Engagement Center. GECIncubator@state.gov

Ghazari-Sargsyan, E., Khoreni-Sargsyan, G., & Inchikyan, H. (1900). *Merdzavor ev Mijin Arevelk'i erkrner ev zhoghovurdner*. Haykakan SSH GA Hratarakch'ut'yun.

Global Disinformation Index. (2020, March 24). *Why is ad tech funding these ads on coronavirus conspiracy sites?* Global Disinformation Index. https://disinformationindex.org/2020/03/why-is-ad-tech-funding-these-ads-on-coronavirus-conspiracy-sites/

Global Voices. (2020, April 19). *How alternative health magazines advance Russia's soft power in the Balkans*. Global Voices. https://globalvoices.org/2019/04/19/how-alternative-health-magazines-advance-russias-soft-power-in-the-balkans/

Government Communications Service. (2019). *RESIST counter-disinformation toolkit*. Government Communications Service.

Grossman, S., Bush, D., & DiResta, R. (2019). *Evidence of Russia-linked influence operations in Africa*. Stanford Internet Observatory.

Gunter, G., & Robinson, O. (2018, September 9). *Sergei Skripal and the Russian disinformation game*. BBC News. https://www.bbc.co.uk/news/world-europe-45454142

Hanna, A. (2020, April 02). *What Islamists are doing and saying on COVID-19 crisis*. Wilson Center. https://www.wilsoncenter.org/article/what-islamists-are-doing-and-saying-COVID-19-crisis

HM Government. (2017). *Internet safety strategy—Green Paper*. HM Government.

HM Government. (2019a). *Online Harms White Paper*. HM Government.

HM Government. (2019b, 1 October). *Privacy notice—Analysis of social media data to tackle disinformation (EU and US partners)*. https://www.gov.uk/government/publications/privacy-notice-analysis-of-social-media-data-to-tackle-disinformation

HM Government. (2020a). *SHARE checklist*. https://sharechecklist.gov.uk/

HM Government. (2020b, March 30). *Government cracks down on spread of false coronavirus information online*. https://www.gov.uk/government/news/government-cracks-down-on-spread-of-false-coronavirus-information-online

HM Government. (2020c, April 15). *Army experts boost NATO fight against COVID-19 disinformation*. https://www.gov.uk/government/news/army-experts-boost-nato-fight-against-COVID-19-disinformation

HM Government. (2021). *Global Britain in a competitive age: The integrated review of security, defence, development and foreign policy*. HM Government.

Hoffman, F. (2018). Examining complex forms of conflict: Gray zone and hybrid challenges. *Prism, 7*(4), 31–47.

Home Office. (2020). *Interim code of practice on terrorist content and activity online*. Home Office.

House of Commons. (2018). *Disinformation and 'fake news': Interim report*. House of Commons.

House of Commons. (2019). *Disinformation and 'fake news': Final report*. House of Commons.

Ibragimova, G. (2020, March 21). *"This is war" Italy lacks virus tests—They are bought by the USA*. RIA. https://ria.ru/20200321/1568907205.html

Innes, M. (2020). Techniques of disinformation: Constructing and communicating "soft facts" after terrorism. *British Journal of Sociology, 71*(2), 284–299. https://doi.org/10.1111/1468-4446.12735

Institute for Development of Freedom of Information. (2015, October 26). *Creeping occupation of Georgia following the 2008 war*. IDFI. https://idfi.ge/en/changed-borders-of-georgia-after-occupation

Jozwiak, R. (2020a, April 27). *NATO's Stoltenberg blasts Chinese, Russian disinformation about coronavirus*. Radio Free Europe/Radio Liberty. https://www.rferl.org/a/nato-s-stoltenberg-blasts-chinese-russian-disinformation-about-coronavirus/30579874.html

Jozwiak, R. (2020b, 23 April). *EU monitors see coordinated COVID-19 disinformation effort by Iran, Russia, China*. RFERL. https://www.rferl.org/a/eu-monitors-sees-coordinated-COVID-19-disinformation-effort-by-iran-russia-china/30570938.html

Kakachia, K. (2018, April). How the West should respond to Russia's "borderization" in Georgia. *PONARS Eurasia Policy Memo, 523*, 1–8.

Kalenský, J. (2020, March 24). *Analysis | Six reasons the Kremlin spreads disinformation about the coronavirus*. Medium.com. https://medium.com/dfrlab/commentary-six-reasons-the-kremlin-spreads-disinformation-about-the-coronavirus-8fee41444f60

Khan, S. (2019, February 5). *Russia's new strategy in Georgia: Creeping occupation*. London School of Economics. https://blogs.lse.ac.uk/humanrights/2019/02/05/russias-new-strategy-in-georgia-creeping-occupation/

Korolev, A. Yu., Koroleva, A. A., & Yakovlev, A. D. (2015). *Maskirovka: Weapons, equipment and objects*. University ITMO.

Koziratsky, Yu. L., Ivantsov, A. V., & Antonovich, P. I. (2015). Aspects of the development of the conceptual apparatus in the field of maskirovka and countering technical intelligence connections. *Military Thought, 2*, 60–71.

Kremidas-Courtney, C. (2019, April 2). *Russian disinformation and the measles surge in Greece*. Ekathimerini.com. https://www.ekathimerini.com/239108/opinion/ekathimerini/comment/russian-disinformation-and-the--measles-surge-in-greece

Kruglanski, A., Gunaratnab, R., Ellenberg, M., & Speckhard, A. (2020). Terrorism in time of the pandemic: Exploiting mayhem. *Global Security: Health, Science and Policy, 5*(1), 121–132. https://doi.org/10.1080/23779497.2020.1832903

Kux, D. (1985). Soviet active measures and disinformation: Overview and assessment. *The US Army War College Quarterly: Parameters, 1*(15), 19–28.

Limonier, L. (2019). *The dissemination of Russian-speaking news media in Africa*. Paris 8 University.

Loomba, S., de Figueiredo, A., Piatek, S., de Graaf, K., & Larson, H. (2021). Measuring the impact of exposure to COVID-19 vaccine misinformation on vaccine intent in the UK and US. *Nature Human Behaviour, 5*, 337–348. https://doi.org/10.1038/s41562-021-01056-1

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied, 27*(1), 1–16. https://doi.org/10.1037/xap0000315

Manning, M., & Romerstein, H. (2004). *Historical dictionary of American propaganda*. Greenwood Press.

Marushchenko, P. N., & Klevtsova, L. A. (2014). Problems of organising maskirovka, aerospace defence objects and ways to solve them. *Military Thought, 8*, 45–50.

Marusic, S. (2020, April 15). *NATO to help North Macedonia combat fake news about virus*. Balkans Insight. https://balkaninsight.com/2020/04/15/nato-to-help-north-macedonia-combat-fake-news-about-virus/

Matveevsky, M. (2017). Military stealth in combat examples. *Army Compilation, 5*, 25–31.

Matsulenko, V. (1975). *The operational camouflage/deception of troops/forces: The experience of the Great Patriotic War*. Military Publishing House.

Mazarr, M. (2015). *Mastering the Gray Zone: Understanding a changing era of conflict*. Strategic Studies Institute and Army War College.

Meek, J. (2020, April 2). *Terrorist groups spin COVID-19 as God's 'smallest soldier' attacking West*. ABC. https://abcnews.go.com/International/terrorist-groups-spin-COVID-19-gods-smallest-soldier/story?id=69930563

MEMRI. (2020, April 13). *ISIS women at Al-Hol refugee camp: Coronavirus does not infect Muslims; only infidels and oppressors die of the virus*. MEMRI. https://www.memri.org/reports/isis-women-al-hol-refugee-camp-coronavirus-does-not-infect-muslims-only-infidels-and

Metodieva, A. (2018, October 2). *How disinformation harmed the referendum in Macedonia*. German Marshall Fund. http://www.gmfus.org/blog/2018/10/02/how-disinformation-harmed-referendum-macedonia

Miller, C., & Colliver, C. (2020). *The 101 of disinformation detection*. Institute for Strategic Dialogue.

Morris, L., Mazaar, M., Hornung, J., Pezard, S., Binnendijk, A., & Kepe, M. (2019). *Gaining competitive advantage in the Gray zone: Response options for coercive aggression below the threshold of major war*. Rand Corporation.

NATO, (2020, April 16). *Russia's top 5 myths about NATO and COVID-19*. YouTube. https://www.youtube.com/watch?v=Zn4z7kCoV8w

Negroponte, N. (1998, December). *Beyond digital*. Wired.com. https://web.media.mit.edu/~nicholas/Wired/WIRED6-12.html

No author. (1927). Roman-Gazeta, page 3 "accurate disinformation is needed" [*nuzhna-tochnaya vyvereniaya dezinformatsiya*]. Goslitizdat, Moscow, USSR.

Ofcom, (2020a). *COVID-19 news and information: Consumption and attitudes*. https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/news-media/coronavirus-news-consumption-attitudes-behaviour

Ofcom. (2020b, July 7). *COVID-19 news and information: Summary of views about misinformation*. Ofcom.

Orlyansky, V. I., & Kuznetsov, N. F. (2013). On the problems of organising operational maskirovka. *Military Thought, 1*, 17–23.

Paul, C., & Matthews, M. (2016). *The Russian "firehose of falsehood" propaganda model—Why it might work and options to counter it*. RAND Corporation.

Pochter, G. (1932). Intelligence service questions. *War and Revolution, 5–6*, 92–110.

Posetti, J., Ireton, C., Wardle, C., Derakhshan, H., Matthews, A., Abu-Fadil, M., Trewinnard, T., Bell, F., & Mantzarlis, A. (2018). *Journalism, fake news and disinformation*. United Nations Educational, Scientific and Cultural Organization.

Powell, L. (2021, January 21). *Video of woman arrested for 'sitting on a bench' was staged, police say*. Microsoft News. https://www.msn.com/en-gb/news/uknews/video-of-woman-arrested-for-sitting-on-a-bench-was-staged-police-say/ar-BB1cCYyT?li=BBoPRmx&amp;ocid=mailsignout

Radio Free Europe/Radio Liberty. (2020, March 30). *Bulgaria charges pro-Russia politician with spreading false information about coronavirus*. https://www.rferl.org/a/bulgaria-charges-pro-russia-politician-with-spreading-false-information-about-coronavirus/30518495.html

Reuters, DPA. (2020, April 25). *Coronavirus song: Filipinos outraged over Chinese embassy's South China Sea 'propaganda video*. https://www.scmp.com/news/asia/southeast-asia/article/3081557/coronavirus-chinese-embassys-one-sea-friendship-video?src=amp-article-related-articles

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on inoculation theory can reduce susceptibility to misinformation across cultures. *The Harvard Kennedy School Misinformation Review, 1*(2), 1–23. https://doi.org/10.37016//mr-2020-008

Roozenbeek, J., & van der Linden, S. (2018). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570–580.

Schultz, R., & Godson, R. (1984). *Dezinformatsia: The strategy of Soviet disinformation*. Brassey's Inc.

Schwartz, M. (2020, June 28). *Cybercrime groups and nation-state attackers blur together*. Bankinfosecurity. https://www.bankinfosecurity.com/cybercrime-groups-nation-state-attackers-blur-together-a-11141

Serov, A. (2011). On the role of disinformation in contemporary conflicts and wars. *Foreign Military Review, 7*, 15–21.

Serov, A. (2019). Disinformation as a foreign policy tool in a number foreign countries. *Foreign Military Review, 8*, 14–20.

Sharwood, S. (2020, March 18). *One whole day: That's how long Facebook's COVID-19 content moderation went without a mess*. The Register. https://www.theregister.co.uk/2020/03/18/facebook_COVID_ai_content_moderation_mistakes/

Shekhovtsov, A. (2020). *Fake election observation as Russia's tool of election interference: The case of AFRIC*. European Platform for Democratic Election.

Silverman, C. (2014). *Verification handbook*. European Journalism Centre.

Statista. (2020, October). *Global digital population as of October 2020*. Statista.com. https://www.statista.com/statistics/617136/digital-population-worldwide/

Taiwan Fact Check Center. (2020, February 27). *False information monitoring warning 2020/2/26*. Taiwan Fact Check Center. https://tfc-taiwan.org.tw/articles/2712

Tass, (2020, April 23). *Russian Foreign Ministry notes reports of coronavirus infection among US forces in Syria*. Tass. https://tass.com/politics/1148861

Thomas, E., & Zhang, A. (2020). *COVID-19 attracts patriotic troll campaigns in support of China's geopolitical interests*. Australian Strategic Policy Institute.

UK Ministry of Defence. (2014). *Joint Doctrine Publication 0-01, UK Defence Doctrine* (5th ed.). London.

UK Ministry of Defence. (2018). *Joint concept note 2/18 information advantage*. MOD.

UK Ministry of Defence. (2020). *Introducing the integrated operating concept*. MOD.

UK Ministry of Defence. (2022, in preparation). *Joint concept note x/22 operating advantage*. MOD.

UK Parliament. (2020, July 21). *Misinformation in the COVID-19 infodemic*. https://publications.parliament.uk/pa/cm5801/cmselect/cmcumeds/234/23404.htm#_idTextAnchor013

United Nations. (2020a, March 23). *Secretary-General's appeal for global ceasefire*. https://www.un.org/sg/en/content/sg/statement/2020-03-23/secretary-generals-appeal-for-global-ceasefire

United Nations. (2020b, April 3). *UN Secretary-General renews appeal for global ceasefire*. https://unric.org/en/un-secretary-general-renews-appeal-for-global-ceasefire/

United Nations. (2020c). *Stop the virus of disinformation*. United Nations Interregional Crime and Justice Research Institute (UNICRI).

U.S. Government. (1987). *Soviet influence activities: A report on active measures and propaganda, 1986–87*. Department of State.

U.S. War Department. (1945). *Russian military dictionary, TM 30-544*. War Department.

van der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifeneder, M. Jaffé, E. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 147–169). Psychology Press.

Verrall, N. (2019). Human behaviour: Big picture thinking. *British Army Review*, BAR Special Report—Culture in Conflict, 114–129. https://www.army.mod.uk/news-and-events/british-army-review/

Verrall, N. (2021). *Future trends and threats from disinformation. DSTL/CP131669*. Presented [online] at the Open Source Communications, Analytic Research (OSCAR) Disinformation Symposium, Cardiff University, 28 April 2021.

Verrall, N., Dunkley, M., Gane, T., & Byrne, R. (2019). Dangerous Liaisons—A 'Big Four' framework that provides a 'HINT' to understanding an adversary's strategy for influence. *RUSI Journal, 164*(3), 52–68. https://doi.org/10.1080/03071847.2019.1643255

Verrall, N., & Mason, D. (2018). The taming of the shrewd—How can the military tackle sophistry, fake news and post-truth in the digital age? *RUSI Journal, 163*(1), 1–9. https://doi.org/10.1080/03071847.2018.1445169

Wardle, C. (2019). *Understanding information disorder*. First Draft.

Weisgerber, M. (2020, March 18). US Air Force flew half a million coronavirus test swabs from Italy to Tennessee. *Defense One*. https://www.defenseone.com/threats/2020/03/us-air-force-flew-half-million-coronavirus-test-kits-italy-tennessee/163879/

White, J. (2016). *Dismiss, distort, distract, and dismay: Continuity and change in Russian disinformation*. Institute for European Studies Policy.

Wong, E., Rosenberg, M., & Barnes, J. (2020, April 22). Chinese agents helped spread messages that sowed virus panic in U.S., officials say. *NY Times*. https://www.nytimes.com/2020/04/22/us/politics/coronavirus-china-disinformation.html

WHO. (2020a, February 15). *Speech at the Munich security conference*. https://www.who.int/dg/speeches/detail/munich-security-conference

WHO. (2020b). *Infodemic management*. https://www.who.int/teams/risk-communication/infodemic-management

Young, G. (2020, April 22). Defence chief says 77th Brigade is countering COVID misinformation. *The National*. https://www.thenational.scot/news/18398012.defence-chief-says-77th-brigade-countering-COVID-misinformation/

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Chapter 5
Web of Lies: Mapping the Narratives, Effects, and Amplifiers of Russian Covid-19 Disinformation

Aiden Hoyle, Thomas Powell, Beatrice Cadet, and Judith van de Kuijt

Abstract The web of lies surrounding COVID-19 has been purposefully exacerbated by hostile actors, with researchers, analysts, and policymakers alike attempting to keep pace with unfolding disinformation narratives and subsequent effects on citizens. While the content of hostile disinformation narratives is relatively well-researched, how these narratives interact and are amplified to generate psychological effects requires further scrutiny. To address this gap, this study uses Russian COVID-19 disinformation combined with network methodologies to contextualize a novel hypothetical model of this process. Specifically, we conduct a content analysis of known disinformation articles about COVID-19 ($N = 65$) from Russian news sources (e.g. RT, Sputnik, New Eastern Outlook). Using co-occurrence network visualizations, we map the nexus between narrative and psychological effects to provide new insights and testable models of the effects of COVID-19 disinformation. Main findings show how hostile anti-Western narratives primarily target the emotions of anger, disgust, and confusion with the aim of undermining citizens' trust in (supra-) governmental institutions and the media. This is the first step in a research agenda that can help media practitioners develop interventions and aid policymakers in bolstering societal resilience to hostile disinformation campaigns.

5.1 Introduction

In a world of increasing digitisation and interconnectivity, the use of technology to facilitate communication and information seeking is, for many, a central part of daily life. This is especially true during the ongoing COVID-19 pandemic, where technology and social media are being widely used to keep people informed and connected (Adhanom-Ghebreyesus, 2020). It is an uncomfortable paradox, however,

A. Hoyle (✉)
University of Amsterdam, Amsterdam, The Netherlands
e-mail: aiden.hoyle@tno.nl

T. Powell · B. Cadet · J. van de Kuijt
Netherlands Organisation for Applied Scientific Research (TNO), The Hague, The Netherlands

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_5

that the freedoms and technology that guarantee the possibilities of being connected and informed also allow malicious actors to engage in disinformation activities that challenge our national security and, in doing so, undermine these freedoms.

Since the outbreak of COVID-19, there has been a worrying trend in the spread of COVID-19 related disinformation and misinformation worldwide, disrupting the global response and jeopardizing measures to control the pandemic. News about the pandemic has dominated social media platforms since the beginning of the crisis, resulting in the dissemination of both useful virus-related information, but also false news, conspiracy theories, magical cures, and racist news (Rathore & Farooq, 2020). This unprecedented proliferation led Dr. Tedros Adhanom-Ghebreyesus, Director-General of the World Health Organization (WHO), to declare an 'info-demic' (Adhanom-Ghebreyesus, 2020), and some analysts to label the spreading of misinformation and disinformation about COVID-19 as faster than that of the actual virus (Barua et al., 2020).

To disperse disinformation about COVID-19, outlets can use one of the most powerful tools of communication campaigns: the narrative (Levinger, 2018). Across the pandemic, stories have been disseminated alleging malign origins to the virus, such as it being created by the American military or French government, that it is being spread by 5G masts or pharmaceutical companies looking to increase medicine sales, or even that it is a punishment by God (Serrato & Wallis, 2020). Anti-West and anti-US narratives have been used in these stories, and they often place Russia as both a victim and saviour. For instance, stories have emphasised the role that Russian aid played in slowing the spread of COVID-19 in Italy (Zakharov & Soshnikov, 2020). These narratives and other COVID-19 rumours are being amplified, incorporating readers' worries, preferences and attitudes and leading them to make deleterious or sometimes dangerous decisions (Barua et al., 2020).

Extant research on disinformation has shown that specific environmental conditions may heighten the opportunity for the spreading of false information and put organizations and citizens at a higher risk of becoming targets for disinformation campaigns (Oh & Adkins, 2018). For example, disinformation campaigns may be more likely to spread in situations where there are high levels of immediate fear or danger, such as after terrorist attacks (Innes et al., 2019). In situations where people experience high degrees of uncertainty, such as health crises, humanitarian emergencies, or natural disasters, people may be more susceptible to misinterpreting and vulnerable to disinformation (Jaiswal et al., 2020; Oh & Adkins, 2018). Unsurprisingly, the COVID-19 pandemic has led to a rising number of reports of psychological distress, which may increase people's susceptibility to disinformation (Colley et al., 2020, p. 99).

However, despite knowledge that hostile disinformation campaigns are underway and legitimate concerns about their impact, currently, there is little knowledge regarding the exact effects of disinformation campaigns and narratives on individuals. Although it is acknowledged that disinformation exploits the emotions of the audience and is linked to declining levels of trust in the media (Svensson, 2020), specific effects of COVID-19 disinformation, and disinformation campaigns in general, constitute relatively under-researched topics (Enders et al., 2020; Hoyle et al., 2021). Yet deepening our knowledge of the specific effects that certain disinformation narratives can elicit in citizens may hold important societal implications, including a more nuanced understanding of the actual damage disinformation causes in societies and opportunities to focus and improve policy and strategic communication endeavours (Colley et al., 2020, p. 126). It is therefore valuable to initiate a research agenda that moves from analysis of the content of disinformation to modelling and testing its subsequent effects and amplifiers. Psychological science, and particularly experimental research, has been outlined as a potentially fruitful avenue for this goal. However, there is a scant amount of existing research into the psychological effects of disinformation campaigns (Colley et al., 2020; Enders et al., 2020; Hoyle et al., 2021). Without research setting out a clear basis for such effects, devising impactful experimental designs is challenging.

The current study sought to rectify this dearth by conducting inductive research into the narratives and targeted psychological effects of COVID-19 disinformation, proposing and contextualising a theoretical model of this process. We also reflect on potential amplifiers that, in the context of the pandemic, may have exacerbated these psychological effects. With these goals in mind, we aim to answer the following research questions: (1) What are the predominant Russian disinformation narratives surrounding COVID-19? (2) What specific emotions and types of psychological trust might be affected by disinformation narratives surrounding COVID-19? (3) What are the amplifiers for disinformation in the context of COVID-19? (4) How can these insights be used to design counter-disinformation campaigns? We hope that the results of this study can provide preliminary insights into the psychological effects of COVID-19 disinformation, which in turn, can contribute to contextualised theoretical models for further empirical validation. By doing so, we aim to equip practitioners with the understanding and analytical approach (van de Kuijt & Powell, 2021) to tackle disinformation by developing countermeasures that address the deleterious societal impact of disinformation and hostile information activities.

5.2 Theoretical Framework: Modelling Disinformation and Its Effects

In this study, we model disinformation—which we define as false information deliberately created with the intent to harm a person, social group, organisation, or country (Wardle & Derakhshan, 2017)—as a process, depicted in Fig. 5.1. Modelling disinformation as a process in this way (see also, Barua et al., 2020; Bouwmeester, 2020) has two main benefits. First, by breaking down its constituent concepts, we can better understand the mechanisms through which disinformation delivers effects on a target audience (Mendling et al., 2007). Second, process modelling supports decision-making about interventions employed to mitigate disinformation, as the focus of any interventions can be specified and then, in turn, be modelled in more detail.

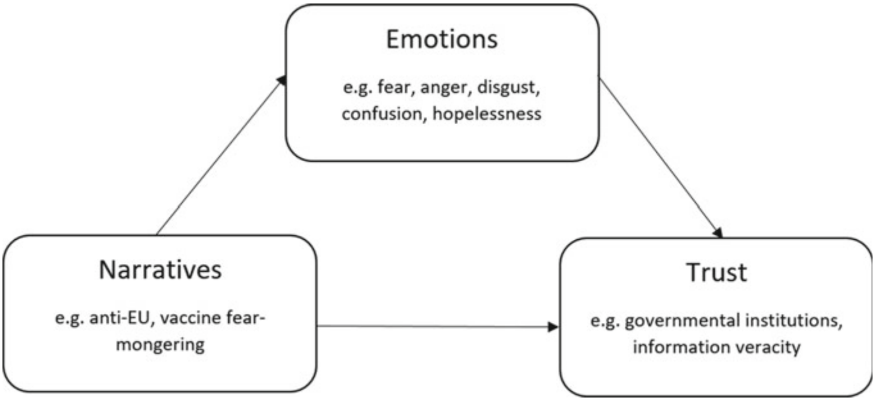

Emotions
e.g. fear, anger, disgust, confusion, hopelessness
Narratives
e.g. anti-EU, vaccine fear-mongering
Trust
e.g. governmental institutions, information veracity

Fig. 5.1 A process (mediation) model of disinformation is advanced in this study (Authors' own image)

The model advanced in this paper draws on concepts from the fields of psychology, communication science, and political science, which together form the theoretical framework of this chapter.

The first element of the model, narratives, are systems of stories *structured* in such a way as to *make meaning* about the world around us (Ruston, 2020). The two terms in italics highlight key elements of narratives. First, narratives consist of content components—known as stories, which are sequences of actions taken by actors in locations—which are structured in a way that they reinforce and complement one another to form a coherent whole, or system. Second, narratives support in making sense and assigning meaning by providing the cognitive building blocks of cause, effect, and consequence required to order information into a mental system, or schema (Rumelhart, 1980). Narratives are different from topics, which refer to shared subject matter, such as COVID-19 (Ruston, 2020). Also different are themes, referring to a dominant or unifying idea, such as 'possible source of COVID-19', without the structural elements of narratives.

Narratives can trigger specific emotions in the receiver. Emotions are neurophysiological states associated with the nervous system, variously related to thoughts, feelings, behavioural responses, and a degree of pleasure or displeasure (Damasio, 1998). Although there is no scientific consensus on a single definition, of particular relevance for COVID-19 disinformation is the psychological theory of emotion called cognitive appraisal theory (Lazarus, 1982; Smith & Ellsworth, 1985). This theory states that emotional responses to stimuli are shaped by how people interpret and assess—in other words, appraise—how the stimuli affect them and their personal situation. Depending on this appraisal, an individual may experience specific emotions—such as fear, anger, disgust—to different extents (Frijda et al., 1989). For example, an anti-US narrative can falsely detail the malign exploitation of the COVID-19 crisis by the US government to exert greater control over the population. A receiver may appraise this as an unfair violation of individual freedoms during a public health crisis, which may trigger the specific emotion of anger towards the government.

The final element of the model, trust, refers to the belief or confidence in the dependability and integrity of someone or something (APA, 2020). In times of crisis, public trust in democracy, authorities, and fellow citizens—dubbed the 'trinity of trusts' (Ingram, 2020, p. 16)—is essential. Particularly so during a global pandemic; citizens need to trust governmental institutions in their policymaking, trust the media in their provision of accurate and timely information, and trust others in society including individual politicians as well as non-governmental institutions to act in the public interest rather than for personal gain. The influence of emotions on the development of (different types of) trust has been widely discussed in discourse around psychological trust (Dunn & Schweitzer, 2005). One can therefore surmise that inflammatory disinformation narratives that evoke negative emotional responses towards these actors (e.g., anger towards institutions, the media, societal groups) may undermine feelings of trust in citizens and precipitate destabilisation in societies.

In Sects. 5.3 and 5.4 that follow, we elaborate on the constituent concepts of this process model in more detail, with a specific attention toward the context of Russian COVID-19 disinformation.

5.3 Russian Master Narratives

We draw on existing research looking at *master narratives* (Levinger, 2018; Rebegea, 2019; Rosen, 2003), and specifically, Russian master narratives. By building on and further developing this work we identify a fuller array of narratives deployed by the Kremlin during the COVID-19 crisis. A master narrative can be defined as a coherent narrative structure, holding meaning and emotional significance based on the history and experiences recognized by a specific community (Levinger, 2018). They are recurring and coherent patterns of narrative use that "provide overarching explanations and justifications for particular stories" (Rebegea, 2019, p. 3). Russian master narratives, therefore, mirror and support the various narratives we can encounter in disinformation campaigns and Russian state-sponsored media outlets. The use of Russia's master narratives relies on the ability to conceal influence under existing tensions and gaps in society and exploit such vulnerabilities.

Master narratives are deployed in support of Russia's strategic goals throughout the region and the broader European continent. In brief, the Kremlin use master narratives in the information domain to distract countries from their European path, deter the adoption of Western values, and participation in current alliance structures (Rebegea, 2019). Master narratives may be intertwined to reinforce each other, or opposite narratives can be deployed to create dissonance and confusion in the reader's mind. This study uses network methodologies to map and visualise the interconnectedness of Russian narratives and unpack how they were used in the COVID-19 crisis.

COVID-19 offers a specific context and ample opportunity to weaponize hostile master narratives. Indeed, the unprecedented challenges for governments on multiple fronts allow for different and multiple angles of attack. For instance, several campaigns have twisted facts to blame European governments for shortfalls in response to the pandemic, playing the *Domestic Failure Master Narrative* (Rebegea, 2019). This narrative attempts to undermine one's trust in their government and has been deployed in combination with the complementary master narrative, *Russia as a Global Player*. This emphasises Russia's successes in stepping in to aid Europe after an opponent has failed to do so. Beyond master narratives, the COVID-19 pandemic also offers novel opportunities to develop new narratives that are tailored to the vulnerabilities of the targeted society and its unfolding situation, termed as *specific narratives* (Rebegea, 2019). Several pressure points surrounding the specific context of the crisis provide opportunities for developing specific narratives. One of the prevalent themes was the misleading campaign referring to the vaccine being artificially created and malignly released from laboratories (Enders et al., 2020). This prominent specific narrative is consistent with the master narratives, most prominently with an *Anti-US* master narrative. As such, specific narratives surrounding COVID-19 can be deployed to reinforce master narratives.

In this study, we assimilate Russian master narratives with narratives specific to COVID-19. This process, described in the Method section, led to the list of Russian COVID-19 narratives shown in Table 5.1 in no particular order. These narratives are then used as a basis for mapping the interconnection between the prominent narratives. In doing so, we show how prominent COVID-19 disinformation narratives reinforce one another to support Russia's strategic goals and messaging.

5.4 (Psychological) Effects of Disinformation

As discussed, there is little literature detailing specific examinations of the psychological effects of disinformation in citizens exposed. However, there are a number of especially pertinent psychological factors one might reason as potential responses to COVID-19 disinformation. The subsequent section identifies two areas of effects, emotions and psychological trust.

5.4.1 Emotions

There is a plethora of research showing that citizens have deep emotional responses to their perceptions of their political or societal reality, and news that outlines and informs these perceptions (Gross & Brewer, 2007; Igartua et al., 2011; Kim & Cameron, 2011). We consider five emotional responses to COVID-19 disinformation: anger, fear, disgust, confusion, and hopelessness. Anger and fear are two emotions that are often posited as responses to perceived threats, something the COVID-19

Table 5.1 Overview of the different narratives coded in the study

Narrative	Description
Anti-EU	This narrative uses the COVID-19 crisis to propagate ideas that the EU had a critically substandard response to the crisis and left certain countries to fend for themselves, or that the crisis is exposing the lack of unity within the bloc
Anti-WHO	This narrative portrays the World Health Organisation as a waning organisation that is purposefully exaggerating the extent of the virus and is under the control of Western governments
Nefarious political elites	The nefarious political elites narrative highlights ways in which shadowy political elites were/could be profiting off the COVID-19 crisis, such as through gaining power or making money off the crisis
Anti-US	The Anti-US narrative situates the United States as a destructive actor in the COVID-19 crisis, either responsible for the creation of the virus, having ulterior intentions in Europe, or using the crisis to attack the Chinese government
Anti-refugees	This narrative leveraged the COVID-19 crisis as a means to demonise refugee communities
COVID-19 data cannot be trusted	This narrative spreads unverified falsities and conspiracy theories about the veracity and trustworthiness of the public information about COVID-19
Fear-mongering about the vaccine	Here, the narrative involves spreading conspiracy theories and rumours about the future COVID-19 vaccines, including that it would include microchips, or that it was being exploited by big-pharma corporations to make money
Pro-Russia or Russia as a victim	This narrative positioned Russia as a saviour of many European states, such as Italy, or propagated the idea that any criticism of Russia is due to inherent Russophobia from the EU and the West that Russia is a constant victim of
COVID-19 is artificially created	This narrative projects the idea that the virus is a biological weapon, which was likely artificially designed in and released from government owned labs—typically owned by the United States—and stationed in Moldova or Ukraine
Anti-NATO	The Anti-NATO narrative centres on NATO's involvement in the crisis, such as their substandard response or their ulterior intentions
COVID-19 is about control	This narrative suggests that the virus is being used as a smokescreen for various governments and organisations to gain control over the World's population
Pro-China	The pro-China narrative focusses on being sympathetic to China, presenting it as unfairly treated by the global community and praising China's quick and stringent response to coronavirus measures

(continued)

Table 5.1 (continued)

Narrative	Description
Anti-West	This narrative centres on denigrating western values, suggesting that Western ways of life were holding back the response to the virus or that unity in the West was slowly crumbling
Anti-Western media	The anti-Western media narrative lambasts the Western media as broadcasters of falsities and inaccuracies, and for their ostensible attempts to rile up panic, instability, and particularly, negative sentiments towards Russia
Impending conflict	This narrative centres on the idea of an impending and unavoidable period of instability, involving the possibility of conflict and the collapse of the international global order, due to changes resulting from the virus

crisis is likely to represent for many citizens (Kachanoff et al., 2020; Lerner & Keltner, 2001). Conceptually, they are closely related emotions, but cognitive appraisal theory delineates their activation across three main dimensions: anger is more likely when a person is certain about a perceived threatening situation, when they can attribute blame to a clear agent, and if they are still able to influence the situation (Lerner & Keltner, 2001). Fear, by contrast, is theorised as a response to ambiguous or indefinable threats that lack clarity regarding responsible parties or the individual's efficacy toward the threat (Rico et al., 2017). In COVID-19 disinformation, cognitive appraisal theory would suggest that criticisms of a specific, elected government's poor decision-making might trigger anger, while allusions to shadowy figures or vague conspiracy theories might trigger fear.

The third emotion we suggest is (moral) disgust, which can be defined as 'a feeling of revulsion elicited by something offensive' (Oaten et al., 2018, p. 1). A growing link is being made between morally aversive stimulants and the emotional experience of disgust, and it has consequently been incorporated in prominent taxonomies of moral emotions (Haidt, 2003). We suggest that disgust can be elicited when one identifies a moral transgression which, in the context of COVID-19 disinformation, might be the portrayal of certain actors causing, benefiting from, or refusing to prevent others' suffering.

We also identify confusion as a relevant emotion.[1] Confusion is defined as the feeling of dissonance that occurs "when individuals encounter incongruence in the form of impasses, anomalies, contradictions, disruptions of goals, extreme novelty that cannot be comprehended" (D'Mello & Graesser, 2014, p. 290). Analysts have highlighted how actors engaging in information operations seek to 'muddy the

[1] There is some debate within emotion psychology regarding whether one can term confusion a genuine emotion (Hutcherson & Gross, 2011) or simply an affective state (Hess, 2003). A broader discussion of this is beyond the scope of this research, and for the purposes of our analysis, we consider it as a valid emotional response to disinformation narratives.

Table 5.2 Overview of the different targeted emotions coded in the study

Emotion	Description
Anger	The emotion of anger was coded when the article projected a well-defined real or imagined threat or injustice that can be accurately attributed to an identifiable culprit who can be punished for their actions
Fear	The emotion of fear was coded when the article projected a real or imagined threat that cannot be blamed on an actor and when the reader has little power to avoid or cope with the negative outcomes of the threatening situation
Disgust	Disgust was coded when the article featured an actor committing a moral transgression, such as deliberately inflicting suffering or ignoring the suffering of others
Confusion	Confusion was coded when an article provided information which likely does not match or is incongruent with the reader's established knowledge or perception of a situation, leading them to question facts and established thoughts
Hopelessness	The emotion of hopelessness was coded when an article sought to undermine one's hope and positive outlook on the future, leading them to believe that a current situation could not lead to a positive outcome

water'—to confuse and disempower citizens by introducing contradictory narratives and information that increases uncertainty (Pomerantsev & Weiss, 2014). This is particularly relevant for COVID-19, being a novel virus, as a focus of the media content will be on disseminating new insights about the virus.

The final emotion suggested is hopelessness, the feeling of despair that one feels when they face a dilemma that cannot be solved (APA, 2020). As discussed, the COVID-19 crisis is likely to constitute a traumatic, life-altering event for most citizens who face widespread and enduring lockdowns, the constant threat of infection, and severe economic consequences. Hopelessness should be triggered when narratives of a lack of progress, help, or optimism for the future are emphasised in media articles (Table 5.2).

5.4.2 Psychological Trust

We also focus on five types of trust: (1) trust in government institutions, (2) trust in politicians or public figures, (3) trust in non-governmental institutions, (4) trust between social groups, and (5) trust in information veracity. Trust, or the loss of trust—arguably increasing in recent years (Hosking, 2019)—is often invoked as a general consequence of hostile information activities (Watts, 2018). Several analysts have attempted to understand how these various elements of trust can be differentially threatened by malign disinformation campaigns (Ingram, 2020).

Of particular relevance to disinformation, and specifically disinformation regarding COVID-19, would be political or institutional trust elements such as trust in the governmental institutions, non-governmental institutions, and in individual politicians. These types of trust broadly describe individuals' evaluations that a (political) institution or actor is credible, fair, transparent, and open (Zmerli, 2014). The undermining of these political and institutional trust factors has been identified as a fundamental aim of many different disinformation campaigns (Colley et al., 2020, p. 118). With the unprecedented spread of the COVID-19 virus thrusting governments into a chaotic crisis where all decision-making was subject to intense scrutiny and contest, hostile actors were presented with a myriad of opportunities to create and disperse inaccuracies regarding their competence and integrity (Roberts, 2020). Non-governmental organisations, such as the World Health Organisation or national medical boards, have also become more prominent in public discourse as COVID-19 spread, conferring new possible targets for disinformation campaigns to disparage.

A further germane element of trust would be trust between the various social groups who co-habit in a society. Undermining interpersonal trust by exploiting contentious issues and scapegoating minority groups has been, similarly to political and institutional trust, recognised as a fundamental goal for (COVID-19) disinformation campaigns (Colley et al., 2020; Flore, 2020, p. 16). Indeed, several early reports noted the emergence of anti-Semitic, sinophobic, and anti-migrant conspiracy theories during the COVID-19 crisis (Fertmann & Kettemann, 2020, p. 11).

A final relevant trust dimension would be trust in information sources. Extant literature regarding disinformation has highlighted a particular method of disinformation campaigns, dubbed 'flooding the zone' (Ramsay & Robertshaw, 2019, p. 6), which describes the extensive proliferation of, often contradicting, narratives in the information environment. This method ostensibly seeks to sew confusion and seeds of doubt in citizens, presenting so many contradictory accounts that citizens disengage and distrust all information sources (Colley et al., 2020, p. 95).

5.5 Methodological Approach

5.5.1 Data Selection and Collection

The study analysed a corpus of 65 articles identified by the EUvsDisinfo project.[2] Outlets logged in this corpus included Russian state-funded media sites such as *RT* or *Sputnik*, and Kremlin-leaning online blogs, and sites such as Geopolitica.ru, Katehon.com, and *New Eastern Outlook*. The period we extracted articles from began on the 16 March 2020 and ended on the 30 April 2020. This was chosen as these dates correspond to the period many conceptualise as making the crucial 'first wave' of COVID-19 in the European continent (Plümper & Neumayer, 2020). The dates, outlets, titles, and content of these articles were then extracted using a web-scraping tool built for this study, and stored in a database. The corpus included only articles

[2] EUvsDisinfo is an open-source data repository coordinated by the European External Action Service's East StratCom Task Force that identifies pro-Kremlin disinformation. https://euvsdisinfo.eu/.

that were written in English and tagged as mentioning COVID-19 by EUvsDisinfo analysts. Prior to the analysis stage, the corpus was checked manually by the authors to remove any articles included by mistake and thus not relevant for the analyses.

5.5.2 Analyses

We adopted a mixed-method approach in this research. The study began with a latent content analysis that was used to distil the different narratives and psychological factors propagated across the corpus generally. Crucially, because of the contextualised nature of the articles, an 'ethnographic content analysis' approach was taken, meaning that when analysing the articles, attention was paid to the 'documentary reality' in which the articles were produced—what Russia's strategic intentions were behind disseminating these narratives (Altheide & Schneider, 2013). As narratives or psychological factors became apparent, they were noted, and as more articles were reviewed, these concepts were refined to fit the corpus. This 'constant comparison' method meant narratives or psychological factors that emerged in articles published later in the selected time period could be identified and integrated into the analysis. This process ended when the coding reached 'saturation' and no new codes were identified in the corpus (Saunders et al., 2018). A taxonomy of concepts was then formalised (Tables 5.1 to 5.3), in which the narratives and psychological factors decided upon were assembled into a coding structure.

We then embraced a more quantitative content analysis approach, whereby the four human coders used the assembled coding structure to review the corpus and code the articles. This method resembles methodologies of studies that examined the narratives propagated in Russian information campaigns, and studies that have investigated emotional reactions to news articles (Gambino & Calvo, 2019; Ramsay & Robertshaw, 2019). The four authors considered the entries before coding a '1' when they felt a narrative was present or a psychological effect was targeted in the article. Hence, the coders were noting presence, rather than strength, of a particular effect. Interrater reliability testing was then conducted to ensure reliability in coding between coders (see Appendix 1). Percentage agreement was high to very high across all categories, and Cohen's Kappa for the whole dataset was 0.68, for narratives was 0.73, for trust factors was 0.7. For the emotions coding, Cohen's Kappa was comparatively lower, averaging 0.52 across the given categories. Due to this falling beneath pre-established rules of thumb for reliability (Lombard et al., 2002), we critically reviewed the coding together before agreeing on the final codes, a method observed (Kluver et al., 2019) and recommended (Lombard et al., 2005) in the field.

The agreed-upon codes were then used for further analysis. This involved looking at the associations between the coded aspects by converting the agreed-upon codes into a co-occurrence matrix. To visualise these associations, the co-occurrence matrix was plotted graphically in a network. A network perspective is a methodological approach that seeks to plot associations between variables as networks. It has been employed in qualitative analysis contexts as a powerful tool to analyse and map

Table 5.3 Overview of the different targeted trust types coded in the study

Type of trust	Description
Trust in government institutions	These reports aimed to damage the impression readers have of governments or institutions, often pushing stories of a lack of control or an impending financial collapse. Other stories aimed to cast unfounded aspersions about the government and their potential exploitation of the crisis to advance other ulterior motives
Trust between societal groups	Articles ascribed with this code aimed to reduce trust and cohesion between different social groups in a society by, for example, suggesting that certain groups were 'responsible' for the spreading of the virus
Trust in non-governmental institutions	These reports aimed to damage the impression readers have of institutions that were not considered governmental, such as medical or media institutions
Trust in politicians/public figures	Entries with this code singled out certain political or public figures as responsible or blameworthy during the crisis. For example, Bill Gates was often singled out as a public figure leveraging the COVID-19 crisis for his personal gain
Trust in information veracity	Disinformation that targeted this factor aims to weaken feelings of certainty in civilians regarding their knowledge of the COVID-19 virus. This can mean the reinforcing of myths and unsubstantiated medical claims regarding who is vulnerable or most affected by the virus, as well as casting general aspersions about its origin and properties

connections between texts across samples of articles (Carley, 1997; Segev, 2020). In a qualitative context, network analysis can consist of nodes representing, for example, words, actors, or concepts, and edges that may represent interactions, co-occurrences, or other types of association. It is also a burgeoning area of psychological science (Schmittmann et al., 2013). In a psychological network analysis, nodes typically represent psychological constructs, such as 'anger' and 'political trust', and edges that indicate statistical relationships between these constructs, such as partial correlations. Analysing and visualising data as networks are seen as fruitful endeavours for those seeking to develop theoretical models and hypotheses about causal relationships (Brugere et al., 2018; Danowski, 1993).

To plot these networks, we used the R-package *qgraph* (Epskamp et al., 2012). We used two different types of networks: one where edges represent the total frequencies of co-occurrence between two coding variables, and a second where edges represent the proportional frequencies of co-occurrence between two coding variables. Both analyses provide different but complementary information. In the total frequency analyses, the edges represent the raw frequency of co-occurrence of two codes across our corpus. The normalised proportional frequency analyses adjust these co-occurrence frequencies to account for the overall number of occurrences of the individual codes. This prevents neglecting co-occurrences that are established in the data, simply because these codes were individually generally infrequent. To calculate this proportional frequency analysis, we converted our agreed-upon codes into similarity coefficients using the 'Jaccard index', yielding a score between 0 and 1, as described in Appendix 2.

5.6 Results

In the following section, we will outline the results of these analyses. The results are graphically depicted in Figs. 5.2 and 5.3. Figure 5.2 displays the total frequencies of each code across the corpus. Figure 5.3 displays four network plots. In all networks, each narrative and psychological factor identified is represented by a node, which is sized depending on their total occurrence in the corpus. Edge-widths correspond to the degree of co-occurrence between the two coding variables.

5.6.1 Narratives

Inspecting Fig. 5.2, and the node sizes of plots (a) and (b) in Fig. 5.3, the most common narratives across the six-week sample were anti-West and anti-US, with over a third of the articles in the corpus identified as projecting these narratives. After these, the next most popular narratives concerned the untrustworthiness of COVID-19 data, that COVID-19 is being used to bolster and extend control of the population, and the anti-EU narrative. Co-occurrence analyses indicated that anti-US rhetoric was frequently associated with the idea of the coronavirus being artificially created by the US (edge 4–9, total co-occurrence [TC] = 12, Jaccard Index [j] = 0.43), stemming from stories that the virus purposefully leaked out of US laboratories. Anti-West rhetoric was frequently associated with Pro-Russia/Russia as a victim (edge 8–13, TC = 11, $j = 0.31$), which in turn was associated with anti-EU rhetoric (edge 13–1, TC = 10, $j = 0.29$). This stems from Russia's projection of how it was a hero in being one of the few to lend help to Western and European countries and how Western media unfairly treats Russia in its coverage (edges 13–14, TC = 9, $j = 0.27$, and 8–14, TC = 6, $j = 0.25$). Anti-West was also associated with narratives implying an impending conflict (edge 13–15, TC = 8, $j = 0.24$), linked to articles detailing how unity in the 'West' was falling apart during the crisis. Other highly associated narratives included stories of nefarious political elites trying to use coronavirus as a vehicle to gain more control (edge 3–11, TC = 11, $j = 0.48$). Plot (b) of Fig. 5.2 indicates that fear-mongering about the vaccine, despite also occurring less frequently across the corpus, was moderately associated with nefarious political elites (edge 3–7, TC = 6, $j = 0.30$), with articles suggesting that the 'elite' are exploiting the vaccine for power or financial gain. Finally, there was also moderate narrative support for China,

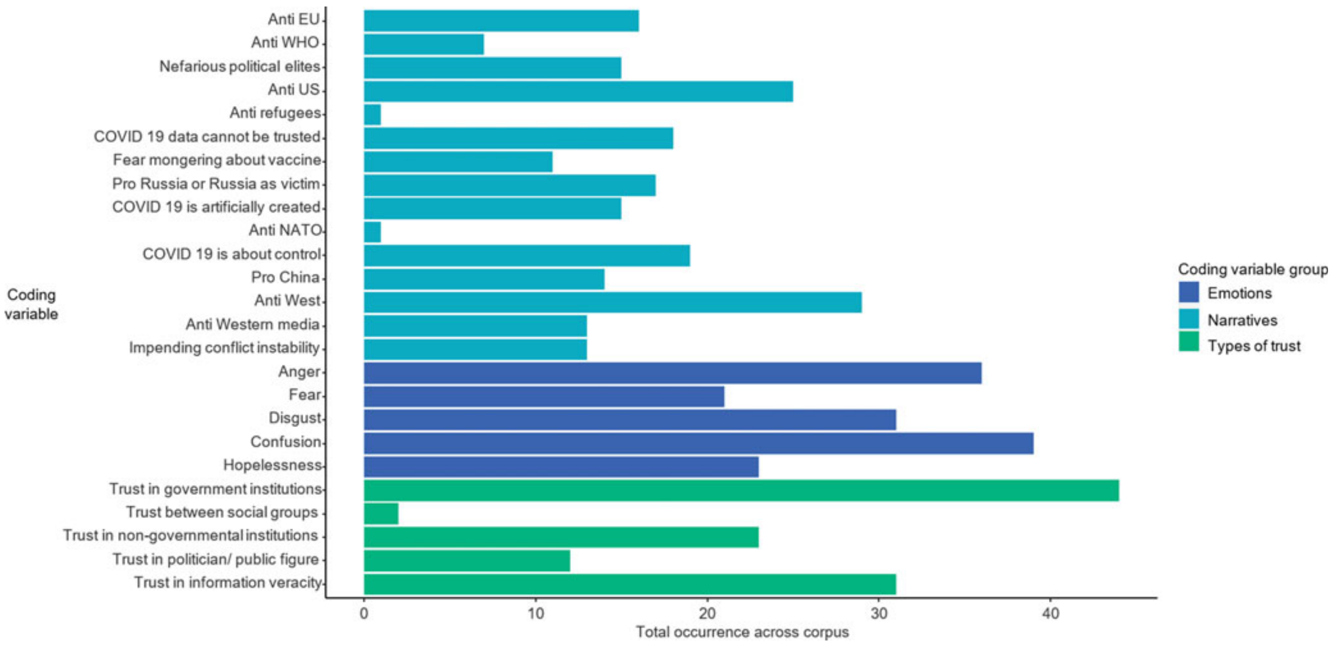


Fig. 5.2 Total occurrences of each coding variable across the corpus

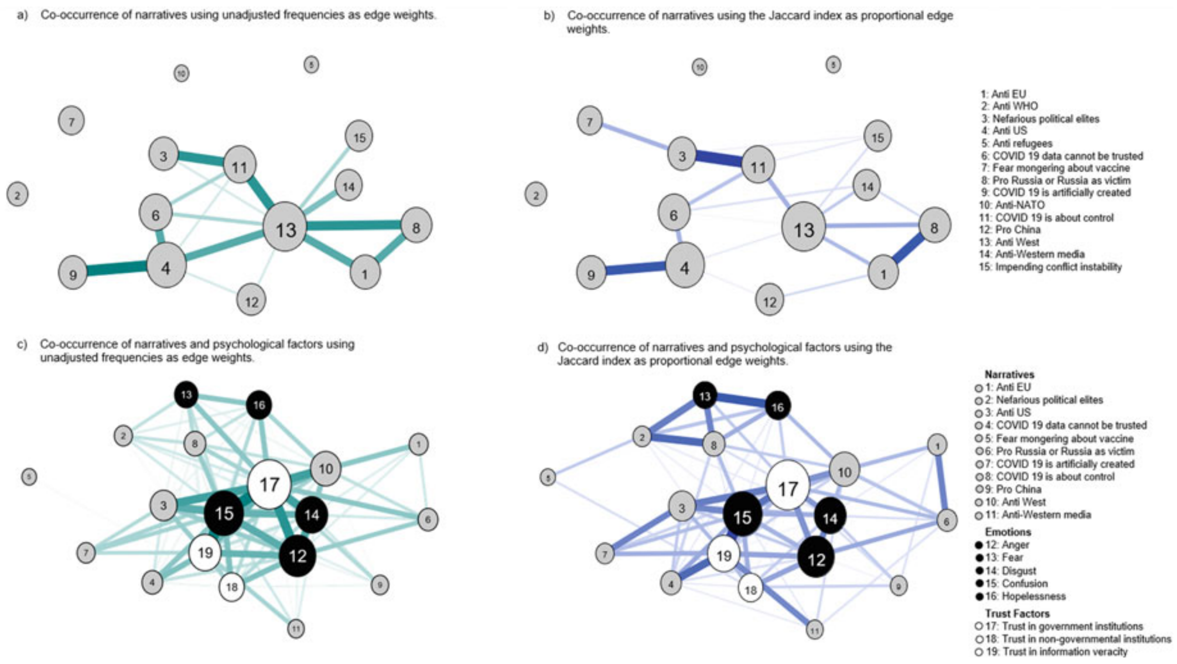

a) Co-occurrence of narratives using unadjusted frequencies as edge weights.
b) Co-occurrence of narratives using the Jaccard index as proportional edge weights.
1: Anti EU
2: Anti WHO
3: Nefarious political elites
4: Anti US
5: Anti refugees
6: COVID 19 data cannot be trusted
7: Fear mongering about vaccine
8: Pro Russia or Russia as victim
9: COVID 19 is artificially created
10: Anti-NATO
11: COVID 19 is about control
12: Pro China
13: Anti West
14: Anti-Western media
15: Impending conflict instability
c) Co-occurrence of narratives and psychological factors using unadjusted frequencies as edge weights.
d) Co-occurrence of narratives and psychological factors using the Jaccard index as proportional edge weights.
Narratives
○ 1: Anti EU
○ 2: Nefarious political elites
○ 3: Anti US
○ 4: COVID 19 data cannot be trusted
○ 5: Fear mongering about vaccine
○ 6: Pro Russia or Russia as victim
○ 7: COVID 19 is artificially created
○ 8: COVID 19 is about control
○ 9: Pro China
○ 10: Anti West
○ 11: Anti-Western media
Emotions
● 12: Anger
● 13: Fear
● 14: Disgust
● 15: Confusion
● 16: Hopelessness
Trust Factors
○ 17: Trust in government institutions
○ 18: Trust in non-governmental institutions
○ 19: Trust in information veracity

Fig. 5.3 Co-occurrence networks showing interactions between the narratives, and between narratives and psychological factors[3]

[3] Plots (a) and (c) use unadjusted frequencies as edge weights (for visualisation purposes, included edges represent a total co-occurrence of > 6). Plots (b) and (d) use Jaccard index as edge weights (for visualisation purposes, included edges represent Jaccard indexes of > 0.20). For similar visualisation purposes, the lowest scoring nodes in plots (c) and (d) have been removed. This means the networks in plot (c) and (d) plot 11 narratives, and eight psychological effects, comprising of five emotions and three types of trust.

which was often juxtaposed against negative coverage of the US (edge 4–12, TC = $7, j = 0.22$) in articles that argued that the success of the Chinese handling was due to their understanding of human rights, an understanding that is 'absent' in the United States.

Taken together, the observed interconnections show the prominence of anti-West and anti-US narratives, reinforced by other master narratives such as Russia as a Victim, and specific narratives such as COVID-19 is artificially created and is about control.

5.6.2 Emotions

The analyses reveal that the most frequently targeted emotions identified across the corpus were anger and confusion, as seen in Fig. 5.2 and inspecting node sizes in plots (c) and (d) of Fig. 5.3. The co-occurrence analyses revealed that these two emotions were also targeted by similar narratives, namely anti-US narratives (edges 3–12, TC = $14, j = 0.30$ and 3–15, TC = $19, j = 0.42$) and anti-West narratives (edges 10–12, TC = $14, j = 0.27$, and 10–15, TC = $17, j = 0.33$). Confusion was also targeted by narratives suggesting that the data around COVID-19 cannot be trusted (edge 4–15, TC = $15, j = 0.36$) and narratives that the COVID-19 virus is artificially created, or being leveraged for control by political actors (edges 7–15, TC = $13, j = 0.32$, and 8–15, TC = $14, j = 0.32$). Anger was also moderately targeted in articles that suggested the COVID-19 data cannot be trusted (edge 4–12, TC = $11, j = 0.26$). Plot (d) indicates that anger's strongest association came from articles that sought to portray Russia as a victim of the West, or that Russia's efforts to help Italy, for example, were being ignored (edge 6–12, TC = $14, j = 0.36$). Disgust was also a highly targeted emotion, again, often stemming from how the conduct of the US (edge 3–14, TC = $14, j = 0.33$) or the West (edge 10–14, TC = $15, j = 0.33$) was portrayed as morally transgressive—using the pandemic for their own gain or withholding information that would reduce their profit. Interestingly, fear was not frequently targeted by the articles in the corpus overall. However, plot (d) shows that when fear was targeted, it was almost entirely conspiracy narratives involving nefarious political elites (edge 2–13, TC = $11, j = 0.44$) or that an elite is leveraging the pandemic to establish greater control in society (edge 8–13, TC = $12, j = 0.43$). Hopelessness was also relatively less frequently targeted, but mainly associated with anti-West narratives that blamed the West's poor handling of the crisis on their individualistic values and lack of unity (edge 10–16, TC = $13, j = 0.33$).

The relations observed show that the two emotions of anger and confusion were most frequently targeted by the prominent anti-West and anti-US narratives, and exacerbated by specific narratives such as COVID-19 data cannot be trusted. Disgust was also a prominently targeted emotion, linked to the moral transgressions of the West, whilst fear was relatively less prominent in the dataset.

5.6.3 Trust Factors

The most targeted trust factors appeared to be citizen's trust in governmental institutions and in the reliability and authenticity of the information about the COVID-19 crisis available to them, as seen in Fig. 5.2 and inspecting node sizes in plots (c) and (d) of Fig. 5.3. Trust in governmental institutions was often associated with narratives that portrayed the United States (edge 3–17, TC = 21, $j = 0.44$) or Western political actors (edge 10–17, TC = 21, $j = 0.40$) in a negative light. It was also highly associated with different emotions, namely anger (edge 12–17, TC = 23, $j = 0.40$), confusion (edge 15–17, TC = 26, $j = 0.46$) and disgust (edge 14–17, TC = 21, $j = 0.39$). Trust in information veracity was also associated with anti-US narratives (edge 3–19, TC = 14, $j = 0.33$), and quite logically, highly associated with narratives that suggested the data about COVID-19 was not trustworthy (edge 4–19, TC = 16, $j = 0.48$). Moreover, trust in information veracity was, again somewhat understandably, highly associated with the targeting of confusion (edge 15–19, TC = 26, $j = 0.59$). Trust in non-governmental institutions was relatively frequently identified, most commonly associated with anti-US narratives (edge 3–18, TC = 8, $j = 0.20$), and the emotions of anger (edge 12–18, TC = 16, $j = 0.37$) and confusion (edge 15–18, TC = 16, $j = 0.35$). Surprisingly, trust between social groups in society and trust in politicians or public figures were not frequently identified as targeted by COVID-19 disinformation in our corpus, identified only 2 and 12 times respectively across the whole corpus, and with very few associations to other variables.

In summary, the most frequently targeted types of trust were trust in government institutions and trust in COVID-19 information, with the co-occurrence network showing these were linked to narratives negatively portraying Western political actors and undermining COVID-19 data.

Specific relationships with targeted emotions were also observed and are mapped in the following section outlining surfacing mediation models.

5.6.4 Surfacing Potential Mediation Models

When our co-occurrence networks are interpreted through the prism of our proposed mediation model of effects, a number of prominent potential mediation models emerge from the data. These models are depicted in Fig. 5.4.

1. *Anti-US*: Firstly, there is a clear link between anti-US narratives and the targeting of confusion and a consequent reduction of trust in governmental structures (edges 3–15–17) and trust in the veracity of information (edges 3–15–19). Similarly, anti-US narratives appear associated with the targeting of anger and citizens' trust in governmental structures (edges 3–12–17).
2. *Anti-West*: Anti-West narratives also show links to anger, disgust, and confusion (edges 10–12, 10–14, and 10–15) which subsequently are associated with effects to governmental and informational trust (–17 and –19). There also seems to be

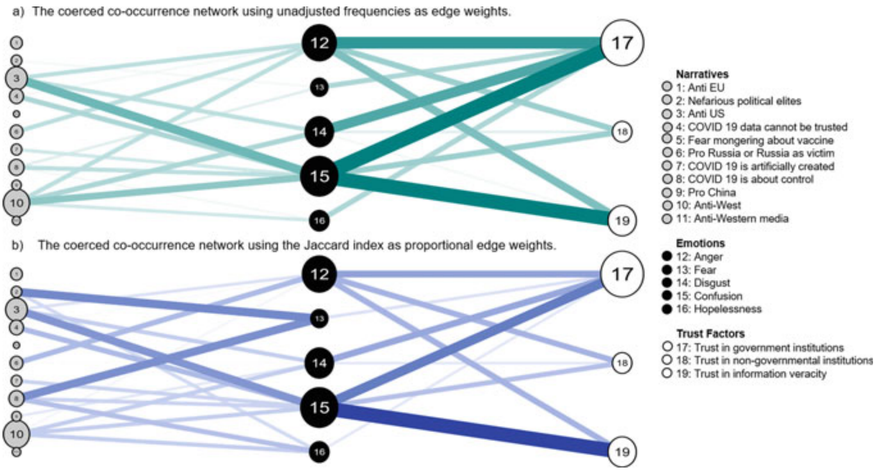

a) The coerced co-occurrence network using unadjusted frequencies as edge weights.
b) The coerced co-occurrence network using the Jaccard index as proportional edge weights.
3
10
12
13
14
15
16
17
18
19
Narratives
○ 1: Anti EU
○ 2: Nefarious political elites
○ 3: Anti US
○ 4: COVID 19 data cannot be trusted
○ 5: Fear mongering about vaccine
○ 6: Pro Russia or Russia as victim
○ 7: COVID 19 is artificially created
○ 8: COVID 19 is about control
○ 9: Pro China
○ 10: Anti-West
○ 11: Anti-Western media
Emotions
● 12: Anger
● 13: Fear
● 14: Disgust
● 15: Confusion
● 16: Hopelessness
Trust Factors
○ 17: Trust in government institutions
○ 18: Trust in non-governmental institutions
○ 19: Trust in information veracity

Fig. 5.4 The co-occurrence analysis networks coerced into the proposed process mediation model structure[4]

a potential mediation between anti-Western narratives, hopelessness, and a loss of trust in governmental institutions (edges 10–16–17).

3. *COVID-19 data cannot be trusted*: Narratives that focus on the untrustworthiness of the published COVID-19 data also appear to trigger confusion, which combined with earlier insights, would appear to suggest a strong link to a reduction in trust of information veracity (edges 4–15–19).
4. *Nefarious political elites & COVID-19 is about control*: Finally, despite low absolute values, narratives that focus on nefarious political elites and that the COVID-19 virus is about control appear strongly associated with the emotion of fear, and subsequently has an (albeit weaker) link to trust in governmental institutions (edges 2–13–17 and 8–13–17).

5.7 Discussion

The current study sought to make the first step in a research agenda investigating the psychological effects of COVID-19 disinformation narratives. We proposed a hypothetical process model of psychological effects, whereby aversive narratives propagated in COVID-19 disinformation incite different emotions in citizens, which in turn, shape the levels of trust citizens have in various aspects of society. To contextualise this proposed hypothetical model, we performed a qualitative content analysis

[4] Plot (a) uses unadjusted frequencies as edge weights (for visualisation purposes, we only include edges representing a total co-occurrence of > 6). Conversely, Plot (b) uses Jaccard index as proportional edge weights (for visualisation purposes, we only include edges representing Jaccard index > 0.20). As before, the lowest scoring nodes have been removed. This means these networks plot 11 narratives, and eight psychological effects, comprising of five emotions and three types of trust.

of the articles captured by EUvsDisinfo, analysing the projected narratives and the trust and emotional factors that these narratives appeared to target. To determine the associations between these narratives and psychological effects, we conducted co-occurrence analyses and plotted these associations as networks. Across the entire corpus, narratives that denigrated and condemned the actions of the United States, and political actors in 'the West' were, in general, the most popular. We identified the emotions of anger and confusion, the trust of governmental institutions, and the trust in the veracity of information as the most frequently targeted psychological factors by Russian COVID-19 disinformation narratives.

In conducting these analyses, the current study confers novel contributions to discourse and research surrounding COVID-19 disinformation, and hostile information activities in general. As mentioned, discussions about the effects of hostile disinformation narratives often lack crucial empirical foundations. With this study, we contribute inductive research that can be used to discuss the effects of COVID-19 disinformation—providing media practitioners and policy makers with important new insights about the sequences of effects different COVID-19 disinformation narratives might trigger. However, in order to make more conclusive statements about the validity of our proposed models, it is imperative that future research verifies these relationships through experimental research. As discussed, experimental research has been advanced as a promising avenue to investigate the (psychological) effects of disinformation narratives, but insights that could be used to create hypotheses that are grounded in the context and reality of disinformation were missing. In surfacing models of the psychological effects targeted by COVID-19 disinformation narratives, the current study makes the first step in this research agenda, offering a plethora of insights that can be readily transposed into workable experimental hypotheses regarding the effects of COVID-19 disinformation.

While this novel method of coding the narratives and targeted psychological factors provides us with valuable insights, it is important to critically reflect on our methodological approach. In particular, initial inter-coder agreement for the emotions category was comparatively lower than that observed for the narrative and trust factors, and fell below commonly agreed thresholds for reliability (Lombard et al., 2002). This points to a difficulty in accurately analysing the targeted emotions, despite all coders using an agreed-upon coding framework. Predicting emotions is something that has been done empirically before (e.g. Gambino & Calvo, 2019), but due to their subjective nature, the identification of emotions targeted in content is more difficult than simply identifying narratives. Although recommended procedures were followed for coding (Lombard et al., 2005) and percentage agreement across the categories was high overall, it does suggest that we should be cautious drawing firm conclusions regarding the emotions. Since this is an exploratory analysis that seeks to contextualise hypothetical models that can be confirmed or falsified later through experimental methods, we believe this limitation does not compromise the findings. It is also worthwhile to mention that studies employing similar binary coding methods to analyse the content of Russian disinformation obtained comparable Cohen's Kappa scores, and have ascribed this to the low proportion of '1' codes across the dataset (Ramsay & Robertshaw, 2019, p. 122).

One might also reflect on the prospects of widening both the time period examined, and the cohort of coders used to analyse the corpus. We examined a period of seven weeks, corresponding to the widely perceived peak of the first wave of COVID-19 on the European continent. Although this choice delivered a wealth of insights about the (dis)information space during COVID-19's emergence, analysing a larger window of time may provide insights about disinformation throughout the pandemic as a whole. Furthermore, a critique of the current study might be that all four coders share a Western European background. Given the wealth of literature examining cultural influences on, for example, emotions (Mesquita, 2003), introducing a wider spectrum of cultural viewpoints might, again, garner insights missed in the current study.

5.7.1 Amplifiers and Vulnerabilities to COVID-19 Disinformation

Discussions of the relationships between narratives and specific psychological effects are incomplete without considering how broader psychological or external factors might amplify these hypothetical relationships. In terms of our hypothetical model, originally presented in Fig. 5.1, these factors would moderate the psychological effects of Russian disinformation narratives (see Fig. 5.5).

The pandemic and the subsequent lockdown in many countries have brought significant changes to one's everyday life. These changes are characterized by an unprecedented level of uncertainty. While most daily activities switched to an almost exclusively online format, in particular on social media, the flow of negative news

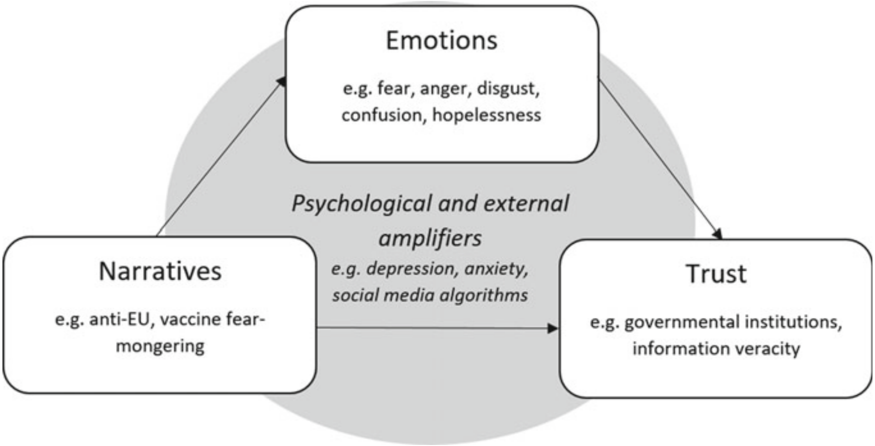

Emotions
e.g. fear, anger, disgust, confusion, hopelessness
Psychological and external amplifiers
e.g. depression, anxiety, social media algorithms
Narratives
e.g. anti-EU, vaccine fear-mongering
Trust
e.g. governmental institutions, information veracity

Fig. 5.5 Psychological and external amplifiers can moderate the effect of disinformation narratives (Authors' own image)

exponentially grew, broadcast by a throng of sources. Meanwhile, opportunities for 'normal' social contact have been limited. As a likely consequence of this situation, there has been an increase in reports of psychological discomfort or distress if not disorders such as depressive symptoms, anxiety, and insomnia (Bu et al., 2020; Pan et al., 2020). These states of psychological distress reported since the start of the COVID-19 crisis are linked with emotional dysregulation (Buckner et al., 2021) which could amplify the effects of disinformation.

Anxiety is characterized as a disruptive feeling of worry or fear, potentially triggered by the uncertainty of a future threat (Grupe & Nitschke, 2013). Depression is itself commonly recognized by symptoms related to hopelessness and sadness, while it has more recently highly been associated with difficulties in fear and anger management (Sahu et al., 2014). Increased levels of anger, fear, sadness, and hopelessness in the general population also mean that citizens approach news media in an elevated state of these negative emotions. Martel et al. (2020) showed that elevated reliance on emotions made subjects more susceptible to disinformation. The intensity of felt emotions was associated with a higher potential to believe articles with disinformation content. Thus, the mental health impact of the pandemic is a likely *psychological amplifier* for the effects of disinformation.

The effects of disinformation may also be aggravated by *external amplifiers*. One prominent example is information overload, both online and offline—the so-called 'infodemic'. As stated above, the pandemic has triggered an exponential growth in the flow of stories surrounding the crisis, with COVID-19 news dominating traditional and social media platforms. The large volume of information means that citizens are presented with more choice in the limited time available for media consumption, which leads to the increased use of mental shortcuts (or heuristics; Chen & Chaiken, 1999) when navigating their media environment. Such mental shortcuts influence the information citizens seek out, comprehend, remember and repeat (Menczer & Hills, 2020) and this can compound biases in information processing (Vosoughi et al., 2018). As such, the sheer amount of information available on the COVID-19 crisis generates immense competition for readers' attention resulting in the loss of high-quality and true information. This plays into the hands of actors seeking to perpetuate falsehoods.

Returning to our hypothetical model, we suggest that psychological and external factors can amplify the effects of disinformation narratives. However, at this stage, we avoid prescribing a specific locus for this moderating effect (see Fig. 5.5). Psychological and external amplifiers may influence the presence and nature of disinformation narratives, as well as how individuals respond to them and how this translates to trust in government and non-government institutions. Moreover, psychological and external amplifiers may interact and reinforce one another. For instance, an increasingly negative information environment surrounding COVID-19 and subsequent negative emotional reactions could motivate people to share content more than usual (Soroka & McAdams, 2015). And the more this content is shared, the more people may feel psychological and emotional distress. More research is needed to unpick such complex effects, as well as to mitigate against them. For instance, future efforts could focus on the effect of mental health care to reduce the impact of conditions like depression and anxiety on emotional regulation. Such efforts may help to bolster resilience against disinformation narratives in times of crisis.

5.7.2 Practical Implications: Counter-Narratives and Building Resilience

A research agenda that validates the observed effects holds promising implications for developing more effects-based countermeasures and policies that address the deleterious societal impact of disinformation and hostile information activities. These implications include more effective construction of counternarratives and the more focussed deployment of societal initiatives. Firstly, an understanding of how propagated disinformation narratives are connected to each other and to different psychological effects means cogent counternarratives can be created that potentially offset, or at least avoid reinforcing, these effects.

Consider, for example, how one might effectively counter anti-US narratives and their effects. Our results section highlighted how a substantial amount of the anti-US narrative stemmed from stories alleging that US laboratories had artificially created the COVID-19 virus. Our analyses further predict that these stories should incite anger, feelings of disgust and confusion in citizens, and foster distrust in governments and available information. These insights, assuming their empirical verification, could valuably inform counter-narrative development as they emphasise the necessity for counter-narrative construction to centre on these emotional and trust factors. In this instance, one might suggest that an effective counter-narrative might disarm anger, confusion, and trust effects by providing clarity to the existence and function of the US laboratories and reassurance of their non-malign intentions. This type of psychological effect-driven counter-narrative construction has already been proposed briefly in counter-terrorism research (Braddock & Horgan, 2016).

We might also consider how such insights can provide evidence to questions regarding what or what not to concentrate societal resilience efforts on. Questions of how to build a society resilient to disinformation are ones that many states are contending with, and indeed many governments are investing large amounts of money and effort into different outreach programs or awareness campaigns in an effort to address such hostile disinformation campaigns (Matasick et al., 2020; Bradshaw & Neudert, 2021). Knowledge of the specific psychological effects that different disinformation campaigns elicit in citizens can inform policymakers about what aspects of the civilian society might be most pressing to reinforce if one wants to effectively counter disinformation campaigns and their impact. Turning to our analyses as an example, we might conclude that the most pertinent aspects of civil society would be citizens' trust in the government and in their information resources. These insights might provide credence to policy proposals that can address these psychological effects, such as open-houses by media organisations, the facilitation of collaboration between news media and fact-checking organisations, and transparent government communication (Devlen, 2020; OECD, 2020). Validation of our results might also suggest that there is relatively less urgency for bolstering resilience in trust between groups in society, or in non-governmental institutions, as these psychological effects did not appear as frequently targeted by the Russian COVID-19 disinformation narratives in our sample.

5.8 Conclusion and Future Research

The current study set out to explore the narratives, potential psychological effects, and potential amplifiers of COVID-19 disinformation. With the psychological effects of disinformation constituting an underdeveloped topic in disinformation discourse, we advanced a hypothetical process model which proposed a mediation relationship between disinformation narratives, emotions elicited by these narratives, and these emotions' effects on (the various types of) psychological trust. To contextualise this model, we conducted content analyses on a sample of articles from pro-Kremlin media which focussed on COVID-19, where we analysed the narratives projected by these articles, and emotions and types of trust we saw as targeted by these articles. We then calculated the co-occurrence between these codes and plotted these associations as networks to generate a contextualised mapping of this proposed model. In inspecting these co-occurrence associations, several prominent pathways along this proposed mediation model were surfaced. We finished by discussing personal and situational factors that may amplify these effects, and possible practical implications if this model is verified. In executing this exploratory research, the study offers low hanging fruit for future research; the demarcated association structures can be easily converted into hypotheses and tested through experimental or survey-based methods. Such endeavours can provide a deeper understanding of the consequences that the COVID-19 'infodemic' has had, and may continue to have, in society. Moreover, it can help in developing countermeasures that address the deleterious societal impact of disinformation and hostile information activities more generally.

Supplementary Material to *Web of Lies: Mapping the Narratives, Effects and Amplifiers of Russian COVID-19 Disinformation*

Appendix 1

Full reliability results for each code in our coding structure.

Variable	Percentage agreement (%)	Cohen's Kappa score	N cases	N agreement	N disagreement
Anti-EU	96.9	0.91	65	63	2
Anti WHO	96.9	0.86	65	63	2
Nefarious political elites	87.7	0.65	65	57	8
Anti-US	86.2	0.70	65	56	9
Anti-refugees	98.5	0.66	65	64	1
COVID-19 data cannot be trusted	84.6	0.58	65	55	10
Fear-mongering about the vaccine	96.9	0.89	65	63	2
Pro-Russia or Russia as a victim	92.3	0.79	65	60	5
COVID-19 is artificially created	92.3	0.79	65	60	5
Anti-NATO	100	1	65	65	0
COVID-19 is about control	87.7	0.67	65	57	8
Pro-China	89.2	0.63	65	58	7
Anti-West	81.5	0.61	65	53	12
Anti-Western media	92.3	0.74	65	60	5
Impending conflict	87.7	0.62	65	57	8
Trust in government institutions	86.2	0.68	65	56	9
Trust between societal groups	98.5	0.79	65	65	1
Trust in non-governmental institutions	78.5	0.52	65	51	14
Trust in politicians/public figures	93.8	0.77	65	61	4
Trust in information veracity	78.5	0.57	65	51	14
Anger	76.9	0.55	65	50	15
Fear	83.1	0.62	65	54	11
Moral disgust	75.4	0.5	65	49	16
Confusion	73.8	0.48	65	48	17
Hopelessness	73.8	0.41	65	48	17

Appendix 2

Calculation of the proportional frequency analysis using Jaccard Index.
The use of the Jaccard Index coefficient is a validated method of normalising co-occurrence data, used in content analysis seeking to identify thematic relations (Patil & Thakur, 2018). It ranges from 0 to 1.0 and can be used to measure how often and consistently two coding variables co-occur. Intuitively, one can interpret Jaccard index as the ratio of the co-occurrence of two coding variable in relation to the total occurrence of either coding variable. For example, a Jaccard index of 0.5 indicates that two coding variables co-occur in 50% of the articles where either of the two is present. It is defined as:

$$j = \frac{n_{12}}{(n_1 + n_2) - n_{12}}$$

where $n1$ = the total number of occurrences of coding variable 1, $n2$ = the total number of occurrences of coding variable 2, and $n12$ = the number of intersecting articles between coding variable 1 and coding variable 2.

References

Adhanom-Ghebreyesus, T. (2020, February 15). *Munich Security Conference*. Retrieved from https://www.who.int/director-general/speeches/detail/munich-security-conference

Altheide, D. L., & Schneider, C. J. (2013). Ethnographic content analysis. In D. L. Altheide & C. J. Schneider (Eds.), *Qualitative media analysis* (pp. 23–37). https://doi.org/10.4135/9781452270043.n2

American Psychological Association. (2020). *APA dictionary of psychology*. American Psychological Association.

Barua, Z., Barua, S., Aktar, S., Kabir, N., & Li, M. (2020). Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Progress in Disaster Science, 8–100119,* 1–9. https://doi.org/10.1016/j.pdisas.2020.100119

Bouwmeester, H. (2020). *Krym Nash: An analysis of modern Russian deception warfare*. Unpublished doctoral thesis, University of Utrecht, Utrecht.

Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict and Terrorism, 39*(5), 381–404. https://doi.org/10.1080/1057610X.2015.1116277

Bradshaw, S., & Neudert, L. M. (2021). *The road ahead: Mapping civil society responses to disinformation*. National Endowment for Democracy's International Forum for Democratic Studies. Retrieved from https://www.ned.org/mapping-civil-society-responses-to-disinformation-international-forum/

Brugere, I., Gallagher, B., & Berger-Wolf, T. Y. (2018). Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys, 51*(2), 1–39. https://doi.org/10.1145/3154524

Bu, F., Steptoe, A., & Fancourt, D. (2020). Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. *Public Health, 186,* 31–34. https://doi.org/10.1016/j.puhe.2020.06.036

Buckner, J. D., Lewis, E. M., Abarno, C. N., Morris, P. E., Glover, N. I., Zvolensky, M. J., & Morris, P. E. (2021). Difficulties with emotion regulation and drinking during the COVID-19 pandemic among undergraduates: The serial mediation of COVID- related distress and drinking to cope with the pandemic. *Cognitive Behaviour Therapy, 50*(4), 1–15. https://doi.org/10.1080/16506073.2020.1861084

Carley, K. M. (1997). Network text analysis: The network position of concepts. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 79–100). Routledge.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). Guilford Press.

Colley, T., Granelli, F., & Althuis, J. (2020). Disinformation's societal impact: Britain, Covid, and beyond. *Defense Strategic Communications, 8,* 89–140. https://doi.org/10.30966/2018.RIGA.8

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Research Reviews, 26*(2–3), 83–86. https://doi.org/10.1016/S0165-0173(97)00064-7

Danowski, J. (1993). Network analysis of message content. In W. D. Richards & G. Barnett (Eds.), *Progress in communication sciences* (12th ed., pp. 198–221). Praeger.

Devine, D., Gaskell, J., Jennings, W., & Stoker, G. (2021). Trust and the coronavirus pandemic: What are the consequences of and for trust? An early review of the literature. *Political Studies Review, 19*(2), 274–285. https://doi.org/10.1177/1478929920948684

Devlen, B. (2020). *How to avoid failure in building societal resilience against disinformation and subversion.* Macdonald-Laurier Institute. Retrieved from https://www.macdonaldlaurier.ca/avoid-failure-building-societal-resilience-disinformation-subversion-policy-brief-balkan-devlen/

D'Mello, S. K., & Graesser, A. C. (2014). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), Educational psychology handbook series. *International handbook of emotions in education* (pp. 289–310). Routledge/Taylor & Francis Group.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology, 88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Enders, A. M., Uscinki, J. E., Klofstad, C., & Stoler, S. (2020). *The different forms of COVID-19 misinformation and their consequences.* Misinformation Review. https://misinforeview.hks.harvard.edu/article/the-different-forms-of-covid-19-misinformation-and-their-consequences/

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1–18. https://doi.org/10.18637/jss.v048.i04

Fertmann, M., & Kettemann, M. C. (2020). *Viral information: How states and platforms deal with COVID-19-related disinformation: An exploratory study of 20 countries* (GDHRNet Working Paper, 1). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI). https://doi.org/10.21241/ssoar.75124

Flore, M. (2020). *Understanding citizens' vulnerabilities (II): From disinformation to hostile narratives.* European Commission Joint Research Centre. https://doi.org/10.2760/271224

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology, 57*(2), 212–228. https://doi.org/10.1037/0022-3514.57.2.212

Gambino, O. J., & Calvo, H. (2019). Predicting emotional reactions to news articles in social networks. *Computer Speech and Language, 58,* 280–303.https://doi.org/10.1016/j.csl.2019.03.004

Gross, K., & Brewer, P. R. (2007). Sore losers: News frames, policy debates, and emotions. *Harvard International Journal of Press/Politics, 12*(1), 122–133. https://doi.org/10.1177/1081180X06297231

Grupe, D., & Nitschke, J. (2013). Uncertainty and anticipation in anxiety. *Nat Rev Neurosci, 14*(7), 488–501. https://doi.org/10.1038/nrn3524.Uncertainty

Haidt, J. (2003). The moral emotions. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford University Press.

Hess, U. (2003). Now you see it, now you don't—The confusing case of confusion as an emotion: Commentary on Rozin and Cohen. *Emotion, 3*(1), 76–80. https://doi.org/10.1037/1528-3542.3.1.76

Hosking, G. (2019). The decline of trust in government. In M. Sasaki (Ed.), *Trust in contemporary society* (pp. 77–103). Brill Publishers.

Hoyle, A., Van den Berg, H., Doosje, B., & Kitzen, M. (2021). Grey matters: Advancing a psychological effects-based approach to countering malign information influence. *New Perspectives, 29*(2), 144–164. https://doi.org/10.1177/2F2336825X21995702

Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100*(4), 719–737. https://doi.org/10.1037/a0022408

Igartua, J. J., Moral-Toranzo, F., & Fernández, I. (2011). Cognitive, attitudinal, and emotional effects of news frame and group cues, on processing news about immigration. *Journal of Media Psychology, 23*(4), 174–185. https://doi.org/10.1027/1864-1105/a000050

Ingram, H. J. (2020). The strategic logic of state and non-state malign 'influence activities': Polarising populations, exploiting the democratic recession. *RUSI Journal, 165*(1), 12–24. https://doi.org/10.1080/03071847.2020.1727156

Innes, M., Dobreva, D., & Innes, H. (2019). Disinformation and digital influencing after terrorism: Spoofing, truthing and social proofing. *Contemporary Social Science, 16*(2), 241–255. https://doi.org/10.1080/21582041.2019.1569714

Jaiswal, J., LoSchiavo, C., & Perlman, D. C. (2020). Disinformation, misinformation and inequality-driven mistrust in the time of COVID-19: Lessons unlearned from AIDS denialism. *AIDS and Behavior, 24*(10), 2776–2780. https://doi.org/10.1007/s10461-020-02925-y

Kachanoff, F. J., Bigman, Y. E., Kapsaskis, K., & Gray, K. (2020). Measuring realistic and symbolic threats of COVID-19 and their unique impacts on well-being and adherence to public health behaviors. *Social Psychological and Personality Science, 12*(5), 603–161. https://doi.org/10.1177/1948550620931634

Kim, H. J., & Cameron, G. T. (2011). Emotions matter in crisis: The role of anger and sadness in the publics' response to crisis news framing and corporate crisis response. *Communication Research, 38*(6), 826–855. https://doi.org/10.1177/0093650210385813

Kluver, R., Cooley, S., & Hinck, R. (2019). Contesting strategic narratives in a global context: The world watches the 2016 U.S. Election. *International Journal of Press/Politics, 24*(1), 92–114. https://doi.org/10.1177/1940161218786426

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American Psychologist, 37*(9), 1019–1024. https://doi.org/10.1037//0003-066x.37.9.1019

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*(1), 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Levinger, M. (2018). Master narratives of disinformation campaigns. *Journal of International Affairs, 71*(1.5), 43–52. Retrieved from https://www.jstor.org/stable/e26508110

Lombard, M., Snyder-duch, J., & Bracken, C. C. (2002). Content analysis in mass communication. *Human Communication Research, 28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

Lombard, M., Snyder-duch, J., & Bracken, C. C. (2005). *Practical resources for assessing and reporting intercoder reliability in content analysis research projects*. Retrieved from http://matthewlombard.com/reliability/

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications, 5*(1). https://doi.org/10.1186/s41235-020-00252-3

Matasick, C., Alfonsi, C., & Bellantoni, A. (2020). Governance responses to disinformation: How open government principles can inform policy options. *OECD Working Papers on Public Governance*, No. 39, OECD Publishing, Paris. https://doi.org/10.1787/d6237c85-en.

Menczer, F., & Hills, T. (2020). Information overload helps fake news spread, and social media knows it. *Scientific American.* Retrieved from https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it/

Mendling, J., Reijers, H. A., & Cardoso, J. (2007). What makes process models understandable? In *Lecture Notes in Computer Science* (pp. 48–63). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75183-0_4

Mesquita, B. (2003). *Emotions as dynamic cultural phenomena.* In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science: Handbook of affective sciences* (pp. 871–890). Oxford University Press.

Oaten, M., Stevenson, R. J., Williams, M. A., Rich, A. N., Butko, M., & Case, T. I. (2018). Moral violations and the experience of disgust and anger. *Frontiers in Behavioral Neuroscience, 12*. https://doi.org/10.3389/fnbeh.2018.00179

OECD. (2020). *Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new coronavirus.* OECD Policy Responses to Coronavirus (COVID-19). Retrieved from https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-bef7ad6e/

Oh, S., & Adkins, T. L. (2018). *Disinformation toolkit*. InterAction. Retrieved from https://www.interaction.org/blog/disinformation-toolkit/

Patil, H., & Thakur, R. S. (2018). Comparative analysis of similarity measures in document clustering. *Information Retrieval and Management*, 47–64. https://doi.org/10.4018/978-1-5225-5191-1.ch003

Penninx, B. W. J. H. (2020). The mental health impact of the COVID-19 pandemic on people with and without depressive, anxiety, or obsessive-compulsive disorders: A longitudinal study of three Dutch case-control cohorts. *The Lancet Psychiatry, 8*(2), 121–129. https://doi.org/10.1016/S2215-0366(20)30491-0

Plümper, T., & Neumayer, E. (2020). Lockdown policies and the dynamics of the first wave of the Sars-CoV-2 pandemic in Europe. *Journal of European Public Policy*, 1–21. https://doi.org/10.1080/13501763.2020.1847170

Pomerantsev, P., & Weiss, M. (2014). *The menace of unreality: How the Kremlin weaponizes information, culture and money*. Institute of Modern Russia.

Ramsay, G., & Robertshaw, S. (2019). *Weaponising news: RT, Sputnik and targeted disinformation.* The Policy Institute: Centre for the study of Media, Communication and Power London. Retrieved from https://www.kcl.ac.uk/policy-institute/assets/weaponising-news.pdf

Rathore, F., & Farooq, F. (2020). Information overload and infodemic in the COVID-19 pandemic. *Journal of the Pakistan Medical Association, 70*(3), 162–165. https://doi.org/10.5455/jpma.38

Rebegea, C. (2019). "Question more"—But not too much: Mapping Russia's malign master narratives in central and eastern Europe. In D. Sultănescu (Ed.), *Challenges in strategic communication and fighting propaganda in eastern Europe* (pp. 75–83). IOS Press. https://doi.org/10.3233/978-1-61499-943-0-75

Rico, G., Guinjoan, M., & Anduiza, E. (2017). The emotional underpinnings of populism: How anger and fear affect populist attitudes. *Swiss Political Science Review, 23*(4), 444–461. https://doi.org/10.1111/spsr.12261

Roberts, H. (2020). Moscow's coronavirus offensive. *Politico.* Retrieved from https://www.politico.eu/article/moscow-coronavirus-offensive-vladimir-putin/

Rosen, J. (2003). *PressThink basics: The master narrative in Journalism* [Blog post]. Retrieved from http://archive.pressthink.org/2003/09/08/basics_master_p.html

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro et al. (Eds.), *Theoretical issues in reading comprehension.* Lawrence Erlbaum.

Ruston, S. (2020). Narrative and strategic communication. In M. Cepurītis, I. Juurvee, A. Keišs, D. Marnot, B. Carrasco Rodríguez, & S. Ruston (Eds.), *Russia's footprint in the Nordic-Baltic information environment* (pp. 6–17). NATO Strategic Communications Centre of Excellence.

Sahu, A., Gupta, P., & Chatterjee, B. (2014). Depression is more than just sadness: A case of excessive anger and its management in depression. *Indian Journal of Psychological Medicine, 36*(1), 77–79. https://doi.org/10.4103/0253-7176.127259

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality and Quantity, 31*(1), 43–53. https://doi.org/10.1007/s11135-017-0574-8

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology, 31,* 43–53. https://doi.org/10.1016/j.newideapsych.2011.02.007

Segev, E. (2020). Textual network analysis: Detecting prevailing themes and biases in international news and social media. *Sociology Compass, 14*(4), 1–14. https://doi.org/10.1111/soc4.12779

Serrato, R., & Wallis, J. (2020). *COVID-19 and the reach of pro-Kremlin messaging*. Australian Strategic Policy Institute's International Cyber Policy Centre. Retrieved from https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-10/Pro%20Kremlin%20messaging.pdf

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Soroka, S., & McAdams, S. (2015). News, politics, and negativity. *Political Communication, 32*(1), 1–22. https://doi.org/10.1080/10584609.2014.881942

Svensson, S. (2020, June 23). *Combatting disinformation: Why trust matters*. International Observatory of Human Rights. Retrieved from https://observatoryihr.org/blog/combatting-disinformation-why-trust-matters/

van de Kuijt, J., & Powell, T. E. (2021). *Am I dealing with disinformation? A guide for signalling disinformation* (in Dutch). Nationaal Coördinator Terrorismebestrijding en Veiligheid. Retrieved from https://www.weerbaar-bestuur.nl/sites/default/files/inline-files/Handreiking%20signaleren%20van%20desinformatie.pdf

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe Report DGI(2017)09. Retrieved from https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Watts, J. T. (2018). Whose truth? Sovereignty, disinformation, and winning the battle of trust. Sovereign challenge conference paper. *Atlantic Council*. Retrieved from https://www.atlanticcouncil.org/in-depth-research-reports/report/whose-truth-sovereignty-disinformation-and-winning-the-battle-of-trust/

Zakharov, A., & Soshnikov, A. (2020). *Russian anthem, Russian flags in Italy, and Pupo singing in Russian: What is wrong with this?* BBC News Russia. Retrieved from https://www.bbc.com/russian/features-52100274

Zmerli, S. (2014). Political trust. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research*. Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_2202

Chapter 6
The Asian Covid-19 Infodemic on Instant Messaging Platforms

Bernice L. Z. Khoo, Shu Jia Chee, and Natalie Lee San Pang

Abstract The Coronavirus Disease of 2019, or COVID-19 epidemic has generated unprecedented levels of uncertainty and confusion regarding public health amongst the world population. A second "pandemic" or *infodemic*, exacerbated by social media, warrants just as much focus as the amplification of COVID-19 misinformation could erode trust, instigate conflict and undermine measures to control the pandemic. This chapter provides a background on the use of instant messaging platforms in Asia and discuss how these platforms are dangerous breeding grounds for misinformation. Through a content analysis of COVID-19 claims circulated via instant messaging platforms, we identified the dominant false narratives, and multimedia (image, video or audio) misinformation. We review the current legislative and non-legislative initiatives that discouraged the creation and propagation of false COVID-19 information as well as the growing body of psychological research on message cues, cognitive vulnerabilities and personality predispositions that make people susceptible to false information. This will contribute toward a more comprehensive strategy to combat misinformation in the longer term.

6.1 Dangerous Breeding Ground for Misinformation

Instant messaging platforms are fast becoming the most popular way to connect with friends, family, acquaintances and businesses. A worldwide survey of the most popular global messenger applications revealed WhatsApp as the top-ranked messenger app with two billion users (*Statistica*, Clement, 2020). This was followed

B. L. Z. Khoo (✉)
DSO National Laboratories, Singapore, Singapore
e-mail: KLinZhi@dso.org.sg

S. J. Chee
DSO National Laboratories, Singapore, Singapore

N. L. S. Pang
National University of Singapore, Singapore, Singapore

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_6

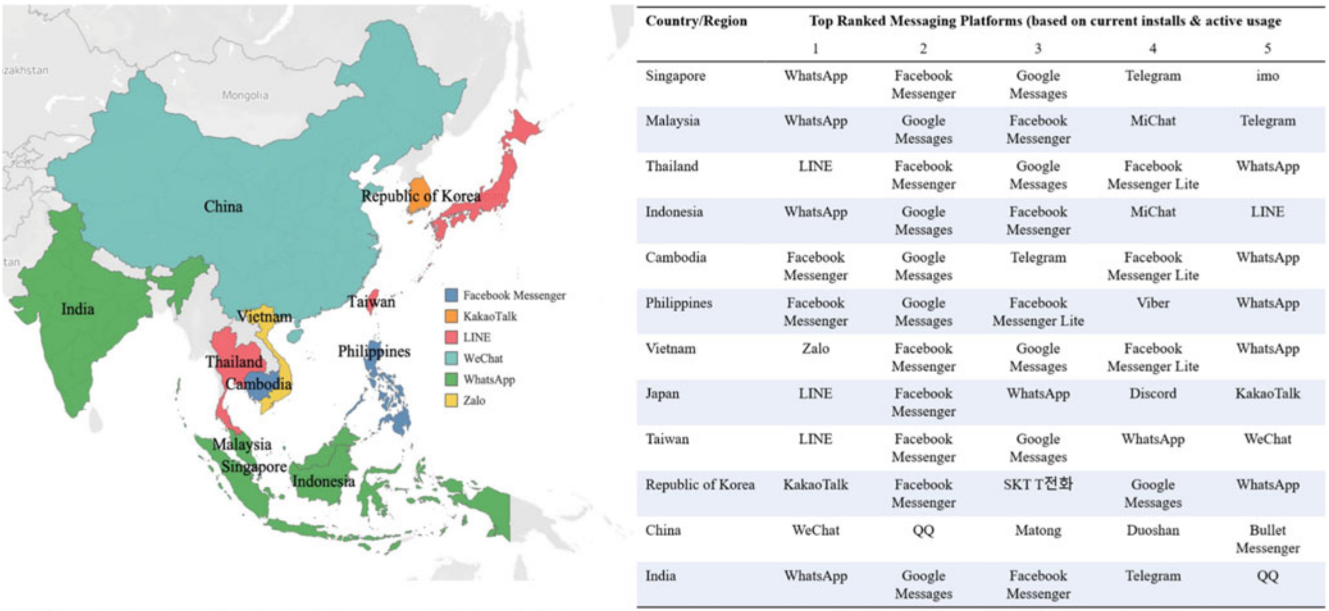
by Facebook Messenger (1.3 billion), WeChat (1.206 billion), QQ (648 million), Snapchat (433 million) and Telegram (400 million). Within Asia, six different applications were identified as the top-ranked messaging platform in eleven countries (Fig. 6.1). They include globally recognized messaging platforms such as WhatsApp, LINE, Facebook Messenger, as well as Zalo, WeChat and Kakao, which have regional monopoly in Vietnam, China, and the Republic of Korea, respectively.

Due to challenges associated with researching instant messaging platforms, mechanisms facilitating misinformation propagation on instant messaging platforms are not well-understood. But recent studies have shown that some features about the communicative environment on messaging platforms make them particularly susceptible as breeding grounds for false information (e.g. Mukhudwana, 2020; Pang & Woo, 2020; Samuels, 2020; Wong & Jensen, 2020).

Instant messaging platforms are distinct when compared to other platforms such as online forums, Facebook and Twitter. Conversations on instant messaging platforms can only be exchanged and seen by members of each chat group. The relatively closed nature of the platform encourages the creation of private groups with members that share many similar views. Exchanges are usually highly reciprocal: Messages sent are often responded to and are intimate in nature to build relationship and social cohesion (Swart et al., 2019). Furthermore, this relatively high degree of trust in each other is also partly driven by growing distrust in media institutions (Pang & Woo, 2020). Since these groups may not be exposed to messages containing divergent views or perspectives, the conditions contribute to the likelihood of groupthink and echo chambers forming within the groups. Although echo chambers do not always generate falsehoods, the likelihood of people believing falsehoods when they are circulated in echo chambers is much higher (Mukhudwana, 2020) due to the high levels of trust and lack of exposure to divergent opinions.

Another reason why instant messaging platforms have grown in popularity is due to the ease of forwarding messages. This enables information to be quickly transmitted between individuals and groups (Wong & Jensen, 2020). This forwarding feature, together with the relatively private and non-public nature of conversations contribute to misinformation and disinformation campaigns increasingly taking root through instant messaging platforms (Bradshaw & Howard, 2018, 2019; Woolley, 2020). Moreover, such campaigns may have dire consequences in real life. For instance, a false video circulating on WhatsApp alleging that there were kidnappers in an Indian village was left unchecked and led to the village believing that the five men handing out chocolates to children were part of a syndicate attempting to kidnap children. This led to a mob attack on these men who were misunderstood as child kidnappers (Samuels, 2020).

Recognising the severity of the problem, the Facebook-owned platform has introduced several measures over the years. For instance, in April 2020, WhatsApp imposed new limits on the number of times a message can be forwarded to slow down the dissemination of falsehoods associated with COVID-19 (Hern, 2020). In August 2020, WhatsApp introduced a pilot fact-check feature to allow users to fact-check



(a) Visual Map of the Top Ranked Messaging Platform in Asia

Country/Region	Top Ranked Messaging Platforms (based on current installs & active usage				
	1	2	3	4	5
Singapore	WhatsApp	Facebook Messenger	Google Messages	Telegram	imo
Malaysia	WhatsApp	Google Messages	Facebook Messenger	MiChat	Telegram
Thailand	LINE	Facebook Messenger	Google Messages	Facebook Messenger Lite	WhatsApp
Indonesia	WhatsApp	Google Messages	Facebook Messenger	MiChat	LINE
Cambodia	Facebook Messenger	Google Messages	Telegram	Facebook Messenger Lite	WhatsApp
Philippines	Facebook Messenger	Google Messages	Facebook Messenger Lite	Viber	WhatsApp
Vietnam	Zalo	Facebook Messenger	Google Messages	Facebook Messenger Lite	WhatsApp
Japan	LINE	Facebook Messenger	WhatsApp	Discord	KakaoTalk
Taiwan	LINE	Facebook Messenger	Google Messages	WhatsApp	WeChat
Republic of Korea	KakaoTalk	Facebook Messenger	SKT T전화	Google Messages	WhatsApp
China	WeChat	QQ	Matong	Duoshan	Bullet Messenger
India	WhatsApp	Google Messages	Facebook Messenger	Telegram	QQ

(b) Top 5 Messaging Platforms in Asia

Fig. 6.1 Instant messaging platforms in Asia (*Source* DeGennaro, 2019; SimilarWeb, 2020)

frequently forwarded messages (Tham, 2020). Still, it is up to users to take the step to actively fact-check information, and the measures taken to combat false information have not been consistently adopted across all instant messaging platforms.

6.2 COVID-19 Misinformation on Messaging Platforms in Asia

To understand the false narratives that were circulated on instant messaging platforms, we identified 131 messages with explicit mention and/or evidence of circulation on one or more instant messaging platforms (e.g. WhatsApp) from 428 online articles on Asian fact-checking websites like Black Dot Research (Singapore), VERA Files (Philippines), Kompas and Turn Back Hoax (Indonesia) and Boom Fake News (India) published between February 2020 and May 2020. Of the 131 messages, 27 messages were from Singapore, 57 messages from Indonesia, 3 messages from Philippines, and 44 messages from India.

False information can take on many different forms. For the purpose of our study, we adopted the following six categories of misinformation and disinformation[1] adapted from Wardle and Derakhshan (2018) and Brennen et al. (2020)[2]:

1. **False context** commonly refers to genuine content that is recirculated out of its original context
2. **Imposter content** comprises content that impersonates authorities and genuine sources of information (e.g. government agencies, established news media) by using their branding
3. **Manipulated content** is genuine content that is intentionally distorted using image, video, or audio editing tools
4. **Fabricated content** refers to content that is fictitious
5. **Misleading content** describes content that has been reframed through misrepresenting or skewing of information (e.g. reframing of headlines, citing of statistics to support different arguments), or presenting opinions and rumours as facts
6. **False connection** happens when a headline or caption does not reflect the content.

Next, we adopted the typology of COVID-19 claims by Brennen et al. (2020) and IDeaS (2020) and identified narratives propagated in the 131 messages. The 8 categories of narratives are in Table 6.1.

[1] The distinction between misinformation, which is the spreading of falsehoods without intent to harm, and disinformation, which includes intent to harm is difficult to infer.

[2] Four categories—misleading content, manipulated content, fabricated content and imposter content were retained from Brennen et al. (2020), and differentiated the types of false context into 2 types—false context and false connection (Wardle & Derakhshan, 2018).

Table 6.1 Eight categories of COVID-19 narratives

Narratives	Example	Clarifications
False emergency or relief measures (e.g. lockdown or quarantine)	"Lockdown in Maharashtra has been extended to April 30"	The decision by the Bombay High Court to extend interim orders till April 30, and was not an extension of the lockdown in Maharashtra (Boom Live, 28 March 2020)
False information on enforcement actions	"Hi guys, random checks are being done by authorities to ensure that everyone in the house are under that particular address"	The Singapore Police Force had come forward to clarify that they "do not proactively conduct checks at residential units" (Black Dot Research, 13 April 2020)
False information on public responses	He told his family to buy food and rice to keep because Singapore is going to shut down…	Although new travel restrictions were introduced at the time that the message was circulated, there was no evidence that Singapore would be implementing lockdown measures yet (Black Dot Research, 17 March 2020)
False information on positive COVID-19 or death	Notification. If possible, don't go to Giant Pal 6 first because the Manager died of Covid and now the employees are being examined	The said Manager had clarified that he had not contracted the virus (Turn Back Hoax, 19 May 2020)
Prevention or treatment measures	Good news. Wuhan's coronavirus can be cured by one bowl of freshly boiled garlic water	Although garlic is considered a common home remedy for common colds, there is no scientific evidence that linked the health benefits of garlic to the treatment of COVID-19 (Boom Live, 1 February 2020)
Virus characteristics	Be careful if you look directly at victims who are infected with the coronavirus, because the coronavirus can be spread by staring. Share as much as possible for good	The virus can only be transmitted through the eyes if an infected person touched his eyes (Turn Back Hoax, 4 February 2020)
Conspiracy theories	…Only doctors and nurses are capable of helping us today. This Corona pandemic has taught me a big lesson that we don't need places of worship…what we need are schools and hospitals. We don't need priests, we need scientists and doctors. I appeal to all the temple priests across the country to shut down the temples forever…	The Rashtriya Swayamsevak Sangh (RSS) functionaries clarified that its Chief, Mohan Bhagwat, did not make such statement (Boom Live, 21 May 2020)

(continued)

Table 6.1 (continued)

Narratives	Example	Clarifications
Others	Scams such as "Free streaming pass from Netflix"	Netflix did not offer any free streaming services as a result of COVID-19 (Boom Live, 25 March 2020)

6.3 Key Findings[3]

The analysis of the categories of misinformation and typology of narratives revealed 3 key findings.

6.3.1 Key Finding 1: Majority of COVID-19 Misinformation on Instant Messaging Platforms Are Reconfigured Content[4]

Our analysis suggests that the predominant technique is to twist or re-contextualize true content to produce half-truths, rather than construct fictitious information. This makes the task of identifying misinformation through verifying claims more complex. 62% of COVID-19 misinformation comprise of *misleading content* such as the reframing of headlines or highlighting one's access to privileged information via family and friends; *false connection* including providing an irrelevant link in the message; *false context*, for instance, the reposting of outdated content; and *manipulated content* such as edited versions of actual infographics (Fig. 6.2).

6.3.2 Key Finding 2: Dominant False Narratives Were Aligned with Real-Time COVID-19 Developments

Different types of narratives gained prominence over the course of the 4 months (Fig. 6.3). In February when the first cases of COVID-19 were reported in Southeast

[3] Fact checked messages from Philippines (VERA Files) ($N = 3$) were excluded from the narrative analysis in Key Findings 2 and 3. Out of 128 fact checked messages, 8 messages (1 from India, 5 from Indonesia and 2 from Singapore) conveyed a second narrative. Therefore, a total of 136 narratives were analyzed in Key Finding 3.

[4] Reconfigured content refers to information that has been edited or recontextualised (Brennen et al., 2020).

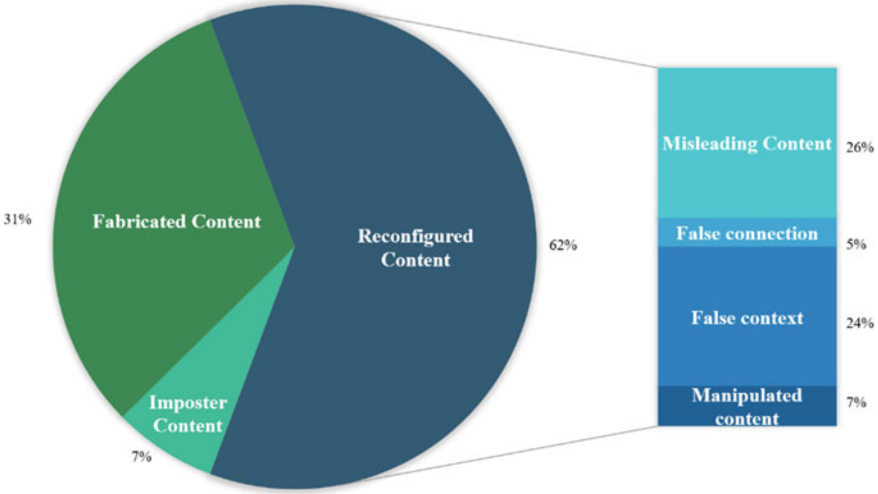


Fig. 6.2 Breakdown of misinformation types and narratives

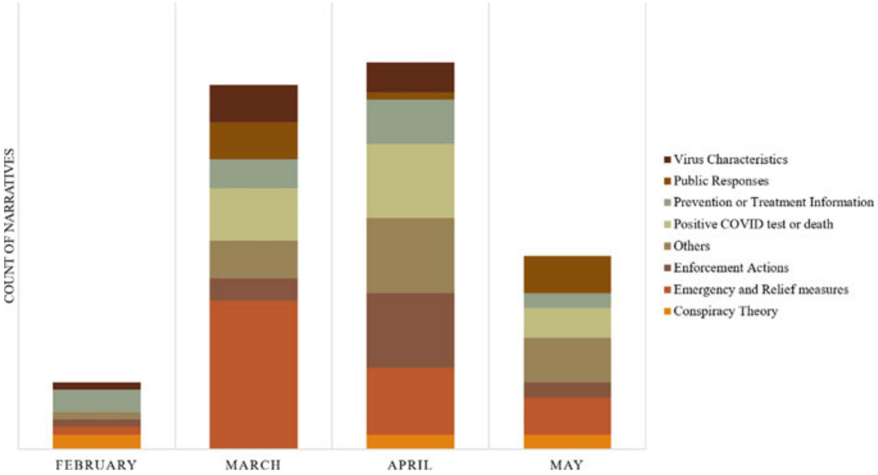


Fig. 6.3 Narrative distribution by month

Asia[5] and little was known about the characteristics of the virus, the majority of false information centered on 'Prevention or Treatment Information', purportedly from authorities or respected sources (e.g. "Ministry of Health", "old Chinese doctor"). As the number of COVID-19 cases climbed in March, there was an exponential increase

[5] The first cases of COVID-19 in Southeast Asia were reported in late January 2020 (Djalante et al., 2020; Fauzi & Paiman, 2020). Singapore, Malaysia, and India reported its first cases on 23, 25, and 27 January respectively. Indonesia reported its first case of COVID-19 on 2 March 2020.

in false information on 'COVID-19 cases' (e.g. "be careful if *(sic)* physical contact with this person, he has tested positive for corona and ran away…"), 'Enforcement Actions' and 'Emergency and Relief Measures' (e.g. nationwide lockdown, closing of schools and workplaces, limits imposed on social gatherings, etc.). These messages contributed to the atmosphere of fear and panic on the lethality of the virus and the need for drastic measures to curb the spread.[6]

As countries introduced their lockdown measures, there were more narratives on 'enforcement actions' in April. In April and May, we observed an increase in entertaining or feel-good stories ('Others') as well as a decrease in false messages on 'Emergency and Relief measures' such as the relaxation of restrictions (e.g. "Tourist attractions in the Special Region of Yogyakarta will open in June 2020") respectively. These narratives appeal to public's desire to return to normalcy.

6.3.3 Key Finding 3: Socio-cultural Elements Are Weaved into False Narratives

While the most common false narrative across all countries was on the implementation of lockdown measures by public authorities ('Emergency and Relief measures'), narratives on 'COVID cases' and 'Prevention or Treatment Information' were more prominent in India and Indonesia than in Singapore (Fig. 6.4). Notably, claims on the efficacy of traditional home remedies (e.g. garlic, lemon) were circulated in both India and Indonesia. Within the 'Others' category, several false Indonesian narratives concerning religious faith were taken out of context and falsely captioned. For instance, a 2011 musical performance in Turkey was recirculated as an interfaith choir chanting *Asmaul Husna* (99 names of Allah) to fight COVID-19. Similarly, a 2018 video of Hindus immersing statues of Lord Ganesh in the sea as part of the Ganesh Chaturthi Festival was recirculated with the accompanying caption: "Hindus in India threw their sculptures into the sea because they did *(sic)* could not defend them from the corona virus. Most Holy Allah, Lord of the Universe".

Compared to India and Indonesia, a greater proportion of false narratives in Singapore were related to the actions and protocols taken by public authorities to enforce the social restriction measures. Since Singapore is a *tight culture* where there is a strong emphasis on social norms (Gelfand, 2018), these false messages leverage on Singaporeans' sensitivity to norm violation to amplify their spread.

[6] By end-March 2020, Singapore, Malaysia, Indonesia and India had only imposed varying degree of social restrictions (e.g. closing schools and workplaces, limiting social gatherings). Full lockdown measures were only implemented later. For example, Singapore implemented the "Circuit Breaker" measures (a stay-at-home order) between 7 April and 1 June, while local governments in Indonesia implemented partial quarantine measures under the *Pembatasan Sosial Berskala Besar*, or PSBB (large-scale social restrictions measures) (CSIS, 2020).

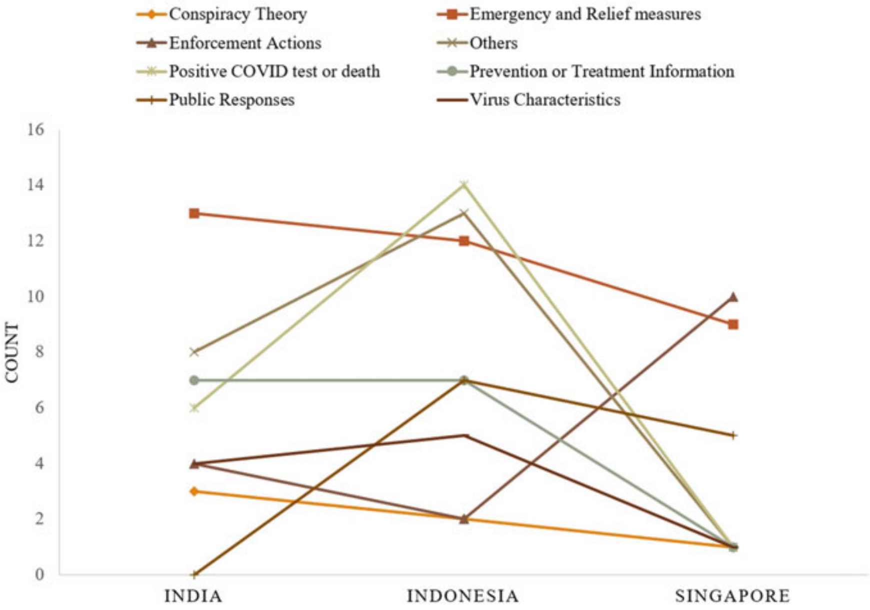


Fig. 6.4 Narrative distribution by country

6.4 Multimedia Misinformation

A significant number of messages in our dataset contained images, audio and videos. Of the 131 messages, 67 contained purely text, 42 contained text with images, audio, video or links, 22 messages contained only images, audio or video only, i.e. no textual information. As shown on Fig. 6.5, the multimedia misinformation in our sample was largely tagged as false context ($N = 19$), followed by misleading content ($N = 15$).

A qualitative analysis was conducted to investigate the functions that visuals served (Brennen, Simon, & Nielsen, 2020). First, visuals served as evidence for claims. For instance, a video of a shipwreck off the coast of Libya in 2014 was not edited or manipulated, but taken out of context along with a claim to induce fear about COVID-19 deaths in India (Boom Live, 7 April 2020). In Philippines, the date on an image of the Enhanced Community Quarantine (ECQ) was manipulated to support the claim the ECQ has been extended to May (VERA Files, 7 May 2020). Second, visuals may be altered to impersonate authorities or authoritative sources to lend credibility to their claims. For instance, a manipulated screenshot of a tweet bearing the logo of ChannelNewsAsia, a media news outlet in Singapore, and the announcement of school closures, was circulated in WhatsApp (Black Dot Research, 7 February 2020). Third, visuals also served to emphasize key aspects of claims, such as excessive enforcement actions against those who violate community guidelines. One such example was a video from a protest rally in Azerbaijan that was shared as

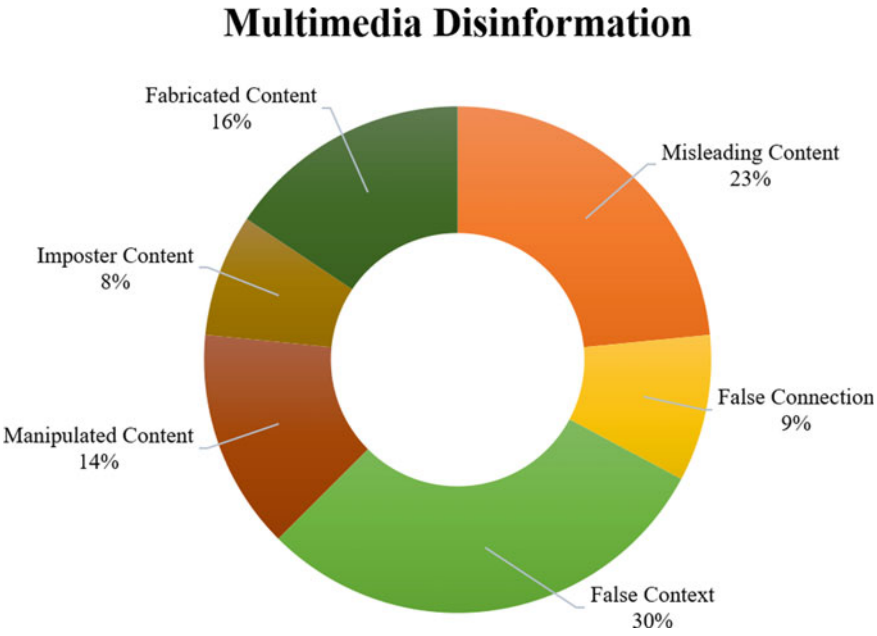


Fig. 6.5 Breakdown of multimedia disinformation

Spanish police enforcing a lockdown, with the caption "*This is lockdown in Spain. You guys in India are lucky..u just get caned*" (Boom Live, 29 March 2020).

Visuals also increased attention, recall and improved comprehension (Houts et al., 2005). Furthermore, Sacchi et al. (2007) found that doctored photographs can also change the way a person remembers past events. For instance, participants who viewed doctored photographs of the 1989 Tiananmen Square protest in Beijing and a 2003 protest in Rome against the war in Iraq experienced memory distortions and rated these events as more violent, negative, confrontational.

Multimodal (text with image) information is often considered more credible textual information (e.g. Hameleers et al., 2020; Hemsley & Snyder, 2018). Nonetheless, on the backdrop that editing technologies are increasingly accessible, visuals such as images or videos should no longer be immediately assumed as reliable. While there has been an overwhelming focus on *deepfakes*—which are videos that have been altered through deep learning algorithms to produce hybrid or synthetic audio-visuals, it would be erroneous to assume that false information must be conveyed using sophisticated technology. One such example is the Nancy Pelosi or "Drunk Pelosi" video, which was slowed to create an impression of slurred speech, or the "drunken" effect (Donovan & Paris, 2019). Based on our analysis as well as similar research by Hameleers et al. (2020), cheapfakes (e.g. slight edits of original images or videos, reusing of old images or videos or the pairing of real visuals with misleading text), still form majority of visual misinformation.

6.5 Approaches to Combat Misinformation

During this pandemic, many countries have explored and implemented their own legislative and non-legislative solutions to combat false information. There is a need to adopt a more holistic approach to deter and counter misinformation. This can be done by incorporating the wealth of scientific research on the conditions and characteristics that predispose individuals towards or buffer them against false information. We discuss some of the predominant psychological theories that identify factors which contribute to individuals being susceptible or impervious to misinformation.

6.6 Legislative and Non-legislative Approaches

Legislative approaches comprise laws and bills to combat misinformation and disinformation. Thailand opened an anti-fake news centre in December 2019 and authorities in Thailand have been working with them to monitor the Internet for falsehoods associated with COVID-19. Based on investigations by Agence France-Presse (AFP) and Philippine Daily Inquirer, people in India, Malaysia, Sri Lanka, Cambodia, Philippines and Indonesia have been fined or arrested for posting COVID-19 related falsehoods (Forster & Feingold, 2020; Marlina, 2020). In Singapore, the Protection from Online Falsehoods and Manipulation Act (POFMA) has been used to issue correction directions for COVID-19 falsehoods circulating on platforms such as online forums, Facebook, WhatsApp and text messaging services (Gov.Sg Website, 2020; Mahmud, 2020).

Non-legislative approaches often combined legislative approaches to combat the spread of false information. Public agencies, private companies and civil society alike have attempted to provide access to quality information so as to combat falsehoods. In April 2020, authorities in Singapore launched new channels on Telegram and WhatsApp which provided daily updates on COVID-19 cases as well as disseminate important announcements associated with the pandemic. Likewise in India and Malaysia, Telegram channels were set up by authorities to broadcast important information and warnings about COVID-19. For example, "Covid19 Karnataka SAHAYA" is a public group operated by the Department of Information and Public Relations in India that enables people to obtain updated information about COVID-19. Tech companies such as Facebook, YouTube and Twitter have also been working to take down false information associated with COVID-19.

Perhaps in response to greater recognition of independent fact-checking bodies, third party fact-checkers have emerged and risen in importance especially in recent years. Many independent fact-checkers operate as subsidiaries or departments associated with media or research companies, and seek endorsement from the International Fact-Checking Network (IFCN), under the Poynter Institute. While they do not need endorsement from IFCN to operate, being endorsed as a fact-checker by IFCN often lends credibility and confidence as fact-checkers that are politically impartial and independent. Tech companies such as Facebook also require fact-checkers that are interested to work with them be endorsed by IFCN. Most IFCN-endorsed fact-checkers are from the United States and Europe, although Asia-based fact-checkers are also emerging. For instance, Boom (India), FactCrescendo (India), JTBC (South Korea), Kompas.com (Indonesia), MyGoPen (Taiwan), Real or Not Myanmar (Myanmar), VERA Files Incorporated (Philippines) are all signatories who have been endorsed by IFCN as fact-checkers in 2020.

6.7 Emerging Psychological Research

In the following section, we discuss the theory of bounded rationality, personality and motivational theories and elaboration likelihood model of persuasion to highlight the role of message cues, cognitive vulnerabilities and personality predispositions in susceptibility to misinformation.

6.7.1 Bounded Rationality and Reliance on Heuristics

The idea of *bounded rationality* (Simon, 1955, 1985) stipulates that humans are not always able to behave rationally due to limitations of mental capacities, knowledge and time (Gigerenzer & Goldstein, 1996; Gigerenzer & Todd, 1999; Kahneman, 2003; Simon, 1982). Because of their limited abilities and resources, people are likely to employ strategies, such as the use of heuristics to reduce cognitive load during information processing. Heuristics serve an evolutionarily important function: helping people cope effectively with the vast quantities of information they encounter every day (Gigerenzer & Todd, 1999). However, in a complex online environment where there are no professional gatekeepers on the online space to engage in quality control over the information that is shared; where information that is shared consists of largely user-generated content and/or lack authority indicators that are crucial for discerning credibility; or when source information may be convoluted by the fact that some information co-produced or aggregated, this reliance on heuristics may inadvertently lead to incorrect conclusions (Metzger & Flanagin, 2013; Metzger et al., 2010).

Specific to instant messaging platforms, false information may be propagated because people trust information from people they know. For instance, O'Keefe (1990) found that, independent of message characteristics such as argument quality, people trust a source that is more familiar compared to an unfamiliar source. Aside from familiarity, other common heuristics such as likability, expertise, or attractiveness of the source have also been used to evaluate the credibility of information (e.g. Petty & Cacioppo, 1996). Moreover, for short messages on instant messaging platforms, information on the credibility of the true source may not be readily available. When this happens, the credibility of the information is a function of the perceived trustworthiness of those who endorse or share the content (Van Der Heide & Lim, 2016).

Aside from whom people receive the information from, cues within the message could impact overall assessments of credibility. To investigate this, we conducted a pilot study to examine the effects of source and calls to actions on credibility ratings of WhatsApp messages. 433 respondents (222 Males, 211 Females), aged between 14 and 26 (Mean = 20.57, SD = 3.18), took part in this study. Respondents were shown screenshots of two or three forwarded WhatsApp messages. The messages were on either of the following three topics: (1) Radiation from regular mobile phone can cause cancer, (2) Chocolates may be an effective remedy for sore throats, and (3) Sleeping through the night is crucial for good health. In addition, each message cited an article that was ostensibly written by either an academic source (e.g. wedmd, Nature, Science) or news media source (e.g. NYTimes, CNN, BBC). We also varied the accompanying message by including a call to action "Do share with your loved ones for the sake of their health" to examine if such accompanying messages affected how respondents perceived the message. After being shown the WhatsApp screenshots, respondents were asked to rate how credible the information was on a scale of 1 (Not credible at all) to 7 (Very credible).

We conducted four sets of analyses to examine the effect of source on credibility ratings when call to action was absent ($N = 233$), the effect of source on credibility ratings when call to action was present ($N = 93$), the effect of call to action on credibility ratings in a message citing an academic source ($N = 93$), and the effect of call to action on credibility ratings in a message citing a news media source ($N = 216$). We hypothesized that respondents would leverage on the *expertise* heuristics, and rate a WhatsApp message that cited an academic source as more credible compared to the same message that cited a media source.

Repeated-measures Analyses of Variance (ANOVA) showed that respondents were more likely to rate a WhatsApp message that cited a media source as more credible than a WhatsApp message that cited an academic source. This effect was unexpected. Seen in a different light, respondents may have adopted the *familiarity* heuristic in their credibility judgments of the WhatsApp message. Since our sample comprises young adults, they may be unfamiliar with academic sources such as wedmd, Nature and Science, and more familiar with new media outlets such as NYTimes, CNN and BBC, therefore they rated the WhatsApp message citing a media source as more credible. We did not find any differences in credibility ratings as a function of call to action.

6.7.2 Personality and Motivation

Aside from a cognitive perspective, susceptibility to misinformation could also be studied from a motivational standpoint. The social identity theory (Tajfel, 1978; Tajfel & Turner, 1979) posits that people want to feel that they are a valuable member of groups. Hence, they are motivated to share new or privileged information as it is a form of social currency that can be used to strengthen social ties and generate an illusion of 'exclusivity' or 'significance' (Tandoc et al., 2020). In a similar vein, Talwar et al. (2019) found that individual traits such as the Fear of Missing Out (FoMO) is positively correlated with self-reported sharing of fake news on WhatsApp. Overall, this suggests that the implicit need to belong drives a tendency to share information including rumours, misinformation or disinformation (Bordia & Difonzo, 2017).

The uses and gratifications approach contend that there are social and psychological motivations that drive people's participation and behaviours on social media, which in turn influences their social media use, interpretations, and sharing of content. Apuke and Omar (2020) found that there were significant association between information seeking and sharing of fake news on the COVID-19 pandemic. In addition, social network site (SNS) dependency, which is one's reliance on SNS as the primary source of information was also associated with fake news sharing. Those high in SNS dependency were more likely to presume online information to be trustworthy and reliable (Lee & Choi, 2018) and share fake news on COVID-19 pandemic (Apuke & Omar, 2020).

Since personality can motivate behaviour, it is also possible that personality characteristics may impact susceptibility to false information. Personality traits such as the Big 5, i.e. openness to experiences, conscientiousness, extraversion, agreeableness and neuroticism, have been shown to impact a diversity of social media behaviours (Gil de Zúñiga et al., 2017). Personality also impacts information seeking, searching and susceptibility to false information. For instance, Wolverton and Stevens (2019) found that people with low openness to experience, extraversion, agreeableness, and conscientiousness and high neuroticism are better able to identify false information. Similarly, other studies found that individuals scoring higher in extraversion accept and disseminate misinformation (Chen & Sin, 2014); individuals with higher conscientiousness are willing to put in more effort when seeking information, and display a preference for respected sources (Heinström, 2003).

To understand the individual differences that predispose an individual to (a) believe that false information is credible (1—Not credible at all to 7—Very credible); (b) unintentionally engage with false information by taking action[7] (1—Very unlikely to 7—Very likely); and (c) intentionally engage with false information by taking action[8] (1—Very unlikely to 7—Very likely), we analysed a subsample of data on the false WhatsApp message of how radiation from mobile phone usage can cause cancer. Respondents ($N = 358$) filled in personality measures, including the Need for Cognition (NfC), Need for Affect (NfA), Empathy Scale, Fear of Missing Out (FoMO), Internet dependency, Purpose of Internet Use, and Big-5 Personality traits.

[7] These actions include the *unintentional* sharing of message with others and changing daily habits according to the message.

[8] These actions include the *intentional* sharing of message with others and changing daily habits according to the message, "just in case".

Overall, findings from our analyses provide preliminary evidence that uses and gratifications-based motivations and personality play influential roles in susceptibility to false information. Consistent with past studies (e.g. Talwar et al., 2019), we found that respondents who scored high on FoMO were more likely to find the message credible and share the message with others. The fear of missing out has been associated with problematic Internet use, such as addictions to social media and video games. At its core, FoMO is driven by a lack of social connections with others and the desire to compensate for such lack can manifest in the ways people use the Internet. This could include being openly supportive of messages as well as sharing or forwarding messages.

In our sample, people who were high in empathy also reported that they would unintentionally and intentionally share false information. This is likely because people who are empathetic may be more concerned about others' well-being, and hence, they may be more likely to share health-related messages with others regardless of message accuracy. In contrast, people who were lower in agreeableness were more likely to report the intentional sharing of false information with others. This could be because they tend to be cynical and have a lower regard for others' interest. Interestingly, respondents who were concerned about self-presentation, were less likely to share any message. This could reflect an inherent status-seeking motivation because the sharing fake messages may have a negative impact on their reputation in the social network (Table 6.2).

Table 6.2 Predictors of susceptibility to false information

Susceptibility	Predictors of susceptibility
Perceive message to be credible	FoMO ($\beta = 0.182$*) Agreeableness ($\beta = 0.125$*)
Action taken: share message with others	FoMO ($\beta = 0.223$**) Empathy ($\beta = 0.188$**) Self-presentation ($\beta = -0.244$***) Informational use ($\beta = -0.116$*)
Action taken: change daily habits accordingly	Relational use ($\beta = 0.152$*) Empathy ($\beta = 0.122$*)
Action taken: intentionally share message with others	Dependency ($\beta = 0.196$**) Empathy ($\beta = 0.172$**) Agreeableness ($\beta = -0.147$*)
Action taken: change daily habits, "just in case"	Empathy ($\beta = 0.179$**) Agreeableness ($\beta = -0.132$*) Internet savviness ($\beta = -0.129$*)

Note Only significant findings are presented *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

6.7.3 Elaboration Likelihood Model: An Information Processing Perspective

Using the elaboration likelihood model of persuasion (ELM; see Petty & Cacioppo, 1981, 1986; Petty & Wegner, 1999) as a theoretical framework, susceptibility to false information can be viewed as a function of motivation and ability factors that impact information processing.

One's ability and motivation to engage in effortful thinking depends on an interplay of individual and situational factors. When motivation and ability are high, people can engage in effortful thinking, and carefully consider information. This is known as the central route. On the other hand, when motivation and/or ability are low, people engage in peripheral processing and tend to assimilate information using cognitive short cuts or heuristics.

Engaging in analytic thinking requires more cognitive effort and resources that may not be readily accessible in the complex online environment where information flows rapidly (Ivaturi et al., 2014). For example, Laato et al. (2020) found that the abundance of novel information on social media during the COVID-19 pandemic led to cognitive overload, which in turn generated social media fatigue and drove users to rely on their intuitive impressions to share unverified information. In the context of ELM, when one's ability to process information is low, their elaboration likelihood is low, and people are more likely to rely on heuristic cues such as clickbait titles and source characteristics rather than on the quality of the arguments (Wiggins, 2017).

Studies also show that younger children and older adults are more susceptible to false information due to lower cognitive ability (Ceci et al., 2007; Wylie et al., 2014). Children are more vulnerable to suggestion because they lack the metamnemonic awareness to shield their memory from suggestion and the awareness to be vigilant about information that may be incongruent (Ceci et al., 2007, 2010). Cognition is also more effortful with age (Hess et al., 2016); older adults are more susceptible to memory errors and source misattributions (Devitt & Schacter, 2016), including false memories for self-relevant information (Rosa & Gutchess, 2013), and loose-binding of source information with own memory (Wylie et al., 2014). When exposed to repeated claims, older adults also tend to over-rely on feelings of fluency when recollection failed them (Skurnik et al., 2005). In the same way, Pennycook and Rand (2018) also showed that people who obtained high scores on the Cognitive Reflection Test were better able to discern fake news and headlines from real news. This effect was observed regardless of whether the news was consistent with one's existing political ideology or not. Taken together, the abovementioned research show that cognitive ability is an important determinant of vulnerability to false information.

Beyond ability, motivational factors can affect whether a person adopts the peripheral or central processing route. One such motivational factor is the Need for Cognition (NfC). Cacioppo et al. (1983) found that individuals high in NfC tend to seek and reflect on the available information, and recall more message arguments regardless of the strength of the message they received. In separate studies, Ferguson et al. (1985), Heppner et al. (1983) and Ahlering and Mcclure (1985) also found that individuals who were higher in NfC engaged in effortful cognitive endeavours such as relying on newspapers and magazines instead of television for news, engaging in personal problem solving and followed presidential debates (Haugtvedy et al., 1988). Of more relevance, Schaewitz et al. (2020) found that one's NfC is an important determinant of accuracy perceptions of information.

Taken together, the dynamics of misinformation propagation is a combination of the medium that is used to transmit the said information, trust of the sender, assessment of the credibility of the information, the recipient's needs, goals and personality, as well as his/her motivation and ability to process information.

6.8 Conclusion

Misinformation is not a new problem; however, it is taking a new turn as instant messaging platforms are fast emerging as hotbeds of misinformation. Moreover, we see that images, audio and videos are increasingly being misused to support false claims and manipulated to create a different reality. In terms of overall counter-misinformation strategy, the current legislative and non-legislative initiatives serve to curb misinformation through discouraging the creation and propagation of false information, i.e. targeting the sender. Since the communication involves both the sender and recipient, more can be done to understand how recipients may be vulnerable to narratives to develop evidence-based solutions to strengthen their capacity and competence to discern false information. To our knowledge, little has been done to develop solutions that specifically combat the rise of multimedia misinformation in Asia. In the United States, for example, the News Provenance Project, is a prominent example of an initiative that aims to develop strategies to circumvent the re-use of old images or videos by labelling the age, location, and original publisher of images. As visuals increasingly become the modality to communicate, it is imperative to start exploring solutions to combat multimedia misinformation and disinformation.

References

Ahlering, R., & McClure, K. (1985). *Need for cognition, attitudes, and the 1984 Presidential election.* Paper presented at the Midwestern Psychological Association Meeting.

Apuke, O. D., & Omar, B. (2020). Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics.* Advance online publication. https://doi.org/10.1016/j.tele.2020.101475

A Singapore Government Agency Website. (2020, April 23). *Clarifications: Misinformation, rumours regarding COVID-19.* Retrieved December 18, 2020, from https://www.gov.sg/article/covid-19-clarifications

Black Dot Research. (2020, February 7). *[COVIDWatch]: Did CNA tweet that schools are closing from 10 Feb?* Retrieved December 13, 2020, from https://blackdotresearch.sg/wuhanwatch-did-cna-tweet-that-schools-are-closing-from-10-feb/

Black Dot Research. (2020, March 17). *[COVIDWatch]: Is Singapore ceasing import of supplies and heading for a lockdown?* Retrieved December 13, 2020, from https://blackdotresearch.sg/covidwatch-is-singapore-ceasing-imports-of-supplies-and-heading-for-a-lockdown/

Black Dot Research. (2020, April 13). *[COVIDWatch]: Are spot checks being done at homes to ensure no social gatherings are happening?* Retrieved December 13, 2020, from https://blackdotresearch.sg/covidwatch-are-spot-checks-being-done-at-homes-to-ensure-no-social-gatherings-are-happening/

Boom Live. (2020, February 1). *Boiled garlic water for treating coronavirus? Not really*. Retrieved December 15, 2020, from https://www.boomlive.in/health/boiled-garlic-water-for-treating-coronavirus-not-really-6737

Boom Live. (2020, March 25). *Scammers, spammers promise free Netflix, Amazon Prime streaming during lockdown.* Retrieved December 15, 2020, from https://www.boomlive.in/fake-news/scammers-spammers-promise-free-netflix-amazon-prime-streaming-during-lockdown-7353

Boom Live. (2020, March 28). *No, The Bombay HC has not extended Maharashtra lockdown till April 30*. Retrieved December 15, 2020, from https://www.boomlive.in/fake-news/no-the-bombay-hc-has-not-extended-maharashtra-lockdown-till-april-30-7414

Boom Live. (2020, March 29). *Video from Azerbaijan falsely shared as Spanish police enforcing a lockdown.* Retrieved December 13, 2020, from https://www.boomlive.in/fake-news/video-from-azerbaijan-falsely-shared-as-spanish-police-enforcing-a-lockdown-7437

Boom Live. (2020, April 7). *Video of migrant shipwreck off Libyan Coast revived with coronavirus spin.* Retrieved December 13, 2020, from https://www.boomlive.in/fake-news/video-of-migrant-shipwreck-off-libyan-coast-revived-with-coronavirus-spin-7585

Boom Live. (2020, May 21). *No, RSS Chief Mohan Bhagwat did not say Covid-19 shook his faith.* Retrieved December 15 2020, from https://www.boomlive.in/fake-news/no-rss-chief-mohan-bhagwat-did-not-say-covid-19-shook-his-faith-8182

Bordia, P., & Difonzo, N. (2017). Psychological motivations. In V. Campion-Vincet, *Rumor mills: The social impact of rumor and legend*. https://doi.org/10.4324/9781315128795-10

Bradshaw, S., & Howard, P. N. (2018). *Challenging truth and trust: A global inventory of organized social media manipulation.* Computational Propaganda Research Project. https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf

Bradshaw, S., & Howard, P. N. (2019). *The global disinformation order: 2019 Global inventory of organised social media manipulation.* Computational Propaganda Research Project. https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). *Types, sources, and claims of COVID-19 misinformation.* Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2020). Beyond (mis)representation: Visuals in COVID-19 misinformation. *The International Journal of Press and Politics*, 1–23. https://doi.org/10.1177/1940161220964780

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45*(4), 805–818. https://doi.org/10.1037/0022-3514.45.4.805

Ceci, S. J., Fitneva, S. A., & Williams, W. M. (2010). Representational constraints on the development of memory and metamemory: A developmental-representational-theory. *Psychological Review, 117*, 464–495. https://doi.org/10.1037/a0019067

Ceci, S. J., Papierno, P. B., & Kulkofsky, S. (2007). Representational constraints on children's suggestibility *Psychological Science, 18*, 503–509. https://doi.org/10.1111/j.1467-9280.2007.01930.x

Center for Informed Democracy & Social-cybersecurity IDeaS. (2020). *List of known misinformation and disinformation regarding corona virus in social media.* https://www.cmu.edu/ideas-social-cybersecurity/research/misinformation-and-disinformation-4-16-2020.pdf

Center for Strategic & International Studies, CSIS. (2020). *Southeast Asia Covid-19 tracker*. Center for Strategic & International Studies. https://www.csis.org/programs/southeast-asia-program/southeast-asia-covid-19-tracker-0

Chen, X., & Sin, S.-C.J. (2014). 'Misinformation? What of it?' Motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology, 50*, 1–4.

Clement, J. (2020, October 29). Most popular global mobile messenger apps as of October 2020, based on number of monthly active users (in millions). *Statistica*. Retrieved November 21, 2020, from https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

DeGennaro, T. (2019, October 16). *The top 3 Chinese chat apps*. Respond.IO. Retrieved November 21, 2020, from https://respond.io/blog/the-top-3-chinese-chat-apps/

Devitt, A. L., & Schacter, D. L. (2016). False memories with age: Neural and cognitive underpinnings. *Neuropsychologia, 91*, 346–359. https://doi.org/10.1016/j.neuropsychologia.2016.08.030

Djalante, R., Nurhidayah, L., Hoang, V. M., Nguyen, T. N. P., Mahendradhata, Y., Trias, A., Lassa, J., & Miller, M. A. (2020). COVID-19 and ASEAN responses: Comparative policy analysis. *Progress in Disaster Science, 8*. Retrieved November 21, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7577870/

Donovan, J., & Paris B. (2019, June 12). Beware the cheapfakes. *Slate*. Retrieved November 21, 2020, from https://slate.com/technology/2019/06/drunk-pelosi-deepfakes-cheapfakes-artificial-intelligence-disinformation.html

Fauzi, M. A., & Paiman, N. (2020). COVID-19 pandemic in Southeast Asia: Intervention and mitigation efforts. *Asian Education and Development Studies*. Ahead of Print. https://doi.org/10.1108/AEDS-04-2020-0064

Ferguson, M., Chung, M., & Weigold, M. (1985). *Need for cognition and the medium dependency components of reliance and exposure*. Paper presented at the International Communication Association Meeting, Honolulu, Hawaii.

Forster, K., & Feingold, S. (2020, April 11). Asia cracks down on virus 'fake news'. *AFP*. Retrieved December 16, 2020, from https://www.straitstimes.com/asia/coronavirus-asia-cracks-down-on-virus-fake-news

Gelfand, M. (2018). *Rule makers, rule breakers: How tight and loose culture wire our world*. Scriber.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & The ABC Research Group, *Evolution and cognition: Simple heuristics that make us smart* (pp. 3–34). Oxford University Press.

Gil de Zúñiga, H., Diehl, T., Huber, B., & Liu, J. (2017). Personality traits and social media use in 20 countries: How personality relates to frequency of social media use, social media news use, and social media use for social interaction. *Cyberpsychology, Behavior, and Social Networking, 20*(9), 540–552. https://doi.org/10.1089/cyber.2017.0295

Hameleers, M., Powell, T. E., Van Der Meer, T. G. L. A., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttal disseminated via social media. *Political Communication, 2*, 281–301. https://doi.org/10.1080/10584609.2019.1674979

Haugtvedt, C., Petty, R. E., Cacioppo, J. T., & Steidley, T. (1988). Personality and ad effectiveness: Exploring the utility of need for cognition. In M. J. Houston (Ed.), *NA—Advances in consumer research* (Vol. 15). Association for Consumer Research.

Heinström, J. (2003). Five personality dimensions and their influence on information behaviour. *Information Research, 9*(1). Retrieved December 18, 2020, from http://www.informationr.net/ir/9-1/paper165.html

Hemsley, J., & Snyder, J. (2018). Dimensions of visual misinformation in the emerging media landscape. In B. Southwell, E. A. Thorson, & L. Sheble (Eds.), *Misinformation and mass audiences*. University of Texas Press.

Heppner, P. P., Reeder, B. L., & Larson, L. M. (1983). Cognitive variables associated with personal problem-solving appraisal: Implications for counseling. *Journal of Counseling Psychology, 30*(4), 537–545. https://doi.org/10.1037/0022-0167.30.4.537

Hern, A. (2020, April 7). WhatsApp to impose new limit on forwarding to fight fake news. *The Guardian*. Retrieved December 10, 2020, from https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news

Hess, T., Matt, C., Benlian, A., & Wiesböck, F. (2016). Options for formulating a digital transformation strategy. *MIS Quarterly Executive, 15*(2), 123–139.

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2005). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counselling, 61*(2), 173–190. https://doi.org/10.1016/j.pec.2005.05.004

Ivaturi, K., Janczewski, L., & Chua, C. (2014). Effect of frame of mind on users' deception detection attitudes and behaviours. In *CONF-IRM 2014 Proceedings, 21*. Retrieved November 23, 2020, from https://aisel.aisnet.org/confirm2014/21

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Laato, S., Islam, A. K. M. N., Islam, M. N., & Whelan, E. (2020). What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? *European Journal of Information System*, 1–18. https://doi.org/10.1080/0960085X.2020.1770632

Lee, J., & Choi, Y. (2018). Informed public against false rumor in the social media era: Focusing on social media dependency. *Telematics and Informatics, 35*(5), 1071–1081. https://doi.org/10.1016/j.tele.2017.12.017

Mahmud, A. H. (2020, October 3). In focus: Has POFMA been effective? A look at the fake news law, 1 year since it kicked in. *ChannelNewsasia*. Retrieved December 18, 2020, from https://www.channelnewsasia.com/news/singapore/singapore-pofma-fake-news-law-1-year-kicked-in-13163404

Marlina, Y. (2020, December 15). Combating fake news in the digital era: Inquirer contributer. *The Straits Times*. Retrieved December 18, 2020, from https://www.straitstimes.com/asia/combating-fake-news-in-the-digital-era-inquirer-contributor

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environment: The use of cognitive heuristics. *Journal of Pragmatics, 59*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility online. *Journal of Communication, 60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Mukhudwana, R. F. (2020). #Zuma must fall this February: Homophily on the echo-chambers of political leaders' Twitter accounts. *Social Media and Elections in Africa, 2*, 175–202. https://doi.org/10.1007/978-3-030-32682-1_10

O'Keefe, D. J. (1990). *Current communication: An advanced text series, Vol. 2. Persuasion: Theory and research.* Sage.

Pang, N., & Woo, Y. T. (2020). What about WhatsApp? A systematic review of WhatsApp and its role in civic and political engagement. *First Monday, 25*(12). https://doi.org/10.5210/fm.v25i12.10417

Pennycook, G., & Rand, D. G. (2018). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition, 188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Petty, R. E., & Cacioppo, J. T. (1981). Issue involvement as a moderator of the effects on attitude of advertising content and context. In K. B. Monroe & A. Abor (Eds.), *NA—Advances in consumer research* (Vol. 8, pp. 20–24). Association for Consumer Research.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). Academic Press.

Petty, R. E., & Cacioppo, J. T. (1996). *Attitudes and persuasion: Classic and contemporary approaches*. Westview Press.

Petty, R. E., & Wegner, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 41–72). Guilford Press.

Rosa, N., & Gutchess, A. H. (2013). False memory in aging resulting from self-referential processing. *The Journals of Gerontology Series B Psychological Sciences and Social Sciences, 68*(6), 882–892. https://doi.org/10.1093/geronb/gbt018

Sacchi, D. L. M., Agnoli, F., & Loftus, E. F. (2007). Changing history: Doctored photographs affect memory for past public events. *Applied Cognitive Psychology, 21*(8), 1005–1022. https://doi.org/10.1002/acp.1394

Samuels, E. (2020, February 21). How misinformation on WhatsApp led to a mob killing in India. *Washington Post*. Retrieved December 13, 2020, from https://www.washingtonpost.com/politics/2020/02/21/how-misinformation-whatsapp-led-deathly-mob-lynching-india/

Schaewitz, L., Kluck, J. P., Klösters, L., & Krämer, N. C. (2020). When is disinformation (in)credible? Experimental findings on message characteristics and individual differences. *Journal of Mass Communication and Society, 23*, 484–509. https://doi.org/10.1080/15205436.2020.1716983

SimilarWeb. (2020, November 11). *Mobile web ranking*. Retrieved November 14, 2020, from https://www.similarweb.com/apps/top/google/store-rank/tw/communication/top-free/

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 6*(1), 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1982). On how to decide what to do. In H. A. Simon, *Models of bounded rationality* (Vol. 2). Behavioral Economics and Business Organization. MIT Press.

Simon, H. A. (1985). Human nature in politics: The dialog of psychology with political science. *The American Political Science Review, 79*(2), 293–304. https://doi.org/10.2307/1956650

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*(4), 713–724. https://doi.org/10.1086/426605

Swart, J., Peters, C., & Broersma, M. (2019). Sharing and discussing news in private socia media groups: The social function of news and current affairs in location-based, work-oriented and leisure-focused communities. *Digital Journalism, 7*(2), 187–205. https://doi.org/10.1080/21670811.2018.1465351

Tajfel, H. (1978). The achievement of inter-group differentiation. In H. Tajfel (Ed.), *Differentiation between social groups* (pp. 77–100). Academic Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of inter-group conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Brooks/Cole.

Talwar, S., Dhir, A., Kaur, P., Zafar, N., & Alrasheedy, M. (2019). Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services, 51*, 72–82. https://doi.org/10.1016/j.jretconser.2019.05.026

Tandoc, E. C., Jr., Lim, D., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism, 21*(3), 381–398. https://doi.org/10.1177/1464884919868325

Tham, I. (2020, August 5). WhatsApp pilots new fact-check feature to spot fake news. *The Straits Times*. Retrieved December 13, 2020, from https://www.straitstimes.com/world/whatsapp-pilots-new-factcheck-feature-to-spot-fake-news

Turn Back Hoax. (2020, February 1). *[Salah] virus corona Dapat Menular Melalui Tatapan Mata*. Retrieved December 15, 2020, from https://turnbackhoax.id/2020/02/04/salah-virus-corona-dapat-menular-melalui-tatapan-mata/

Turn Back Hoax. (2020, May 19). *[Salah] manager giant Pal 6 Banjarmasin Meninggal Kena Covid-19*. Retrieved December 15, 2020, from https://turnbackhoax.id/2020/05/19/salah-manager-giant-pal-6-banjarmasin-meninggal-kena-covid-19/

Van Der Heide, B., & Lim, Y. (2016). On the conditional cueing of credibility heuristics: The case of online influence. *Communication Research, 43*(5), 772–693. https://doi.org/10.1177/0093650214565915

VERA Files. (2020, May 7). *VERA files fact check: Graphic on 'extended' ECQ until May 25 FAKE*. Retrieved December 13, 2020 from https://verafiles.org/articles/vera-files-fact-check-graphic-extended-ecq-until-may-25-fake

Wardle, C., & Derakhshan, H. (2018). Thinking about 'information disorder': Formats of misinformation, disinformation and mal-information. In C. Ireton & J. Posetti, *Journalism, "fake news" & disinformation* (pp. 43–54). UNESCO.

Wiggins, S. (2017). *Discursive psychology: Theory, method and applications*. Sage.

Wolverton, C., & Stevens, D. (2019). The impact of personality in recognizing disinformation. *Online Information Review*. ahead-of-print. https://doi.org/10.1108/OIR-04-2019-0115

Wong, C. M. L., & Jensen, O. (2020). The paradox of trust: Perceived risk and public compliance during the COVID-19 pandemic in Singapore. *Journal of Risk Research, 23*, 1021–1030. https://doi.org/10.1080/13669877.2020.1756386

Woolley, S. (2020, May 1). Encrypted messaging apps are the future of propaganda. *Brookings*. Retrieved November 30, 2020, from https://www.brookings.edu/techstream/encrypted-messaging-apps-are-the-future-of-propaganda/

Wylie, L. E., Patihis, L., McCuller, L. L., Davis, D., Brank, E., Loftus, E. F., & Bornstein, B. (2014). Misinformation effect in older versus younger adults: A meta-analysis and review. In M. P. Tolglia, D. F. Ross, J. Pozzulo, & E. Pica (Eds.), *The elderly eyewitness in court* (pp. 38–66). Psychology Press.

Chapter 7
How to Defend Against Covid Related Disinformation

Jakub Kalenský

Abstract The COVID-19 pandemic has reconfirmed an understanding that has already been recognized over the years: a systematic and well-organized disinformation campaign can threaten peoples' lives, health, and well-being. The leading information aggressors, among which the most experienced is the Kremlin, have significant knowledge and skills aimed at conducting disinformation campaigns very effectively. This chapter describes various defensive countermeasures that can be applied against disinformation, falling into four overarching categories: (1) Documenting the threat; (2) Raising awareness about the threat; (3) Repairing the weaknesses within the information system; (4) Punishing the information aggressors. These categories focus on the desired outcome of the countermeasure, rather than the actor(s) intended to implement them, so as to account for the differences among various countries in the Euro-Atlantic space. Disinformation surrounding the COVID-19 pandemic should not be viewed as an isolated phenomenon, but rather as a campaign that builds on significant disinformation efforts by the same actors many years prior. Countering disinformation needs a more holistic approach, as it requires significantly higher devotion and determination from the West. Once the West decides to begin to effectively deal with the problem, the necessary knowledge and toolbox for action are available.

Author's Note: I would like to thank Teah Pelechaty, Research Assistant in the European Values Center, for her review of the text. I also thank several government officials who wished to remain anonymous for their comments on the text. All views and mistakes are entirely mine.

J. Kalenský (✉)
Atlantic Council's Digital Forensic Research Laboratory, Washington, DC, USA
e-mail: jakub.kalensky@hybridcoe.fi

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_7

7.1 Introduction

The coronavirus pandemic (COVID-19) has shown on a global scale that disinformation can threaten peoples' lives. There have been many incidents, even before the pandemic, that have clearly indicated how dangerous systematic disinformation can be. In 2016, Edgar Maddison Welch drove from North Carolina to a Washington, DC, pizzeria, where he fired a military-style assault rifle into the walls, a computer desk and a door, under the false belief that he was saving children trapped in a sex-slave ring (Haag & Salam, 2017). Czech pensioner, Jaromír Balda, was heavily influenced by years-long consumption of media within an ongoing disinformation campaign denigrating Muslims, which led him to be concerned that the degree of attention paid to this topic in the Czech Republic was insufficient. In 2017, Balda staged terror attacks falsely attributed to Muslims, and tried to derail trains on a local railway. By killing, or simply injuring, innocent people, he had hoped to incite more hatred against Muslim migrants (BBCNews, 2019a). Anti-vaccine activists have been "terrorizing doctors into silence" years before the COVID-19 pandemic started (Karlamangla, 2019). This disinformation campaign has unsurprisingly received a boost from the activity of Russian trolls (O'Kane, 2019). Some researchers would argue that the Kremlin's disinformation campaigns are directly responsible for altered election results worldwide, as in the case of the US 2016 elections (Mayer, 2018). As well, the cases in which disinformation has directly led to protests, chaos and violence against governments are numerous (Barnes & Goldman, 2020; EUvsDisinfo, 2020a; McGuiness, 2016).

However, many of these cases have been limited both temporally and territorially, and the pattern of systematic disinformation campaigns may easily be missed by those who follow only singular events in their respective countries, or who do not feel the need to look at the big picture. Systematic disinformation campaigns regularly threaten not only our freedom of choice, but also peoples' lives and well-being. The tendency of those within many Western democracies to not appreciate the whole picture, or to see this menacing pattern, is further bolstered by the Kremlin's disinformation channels, some within the academic community (Galeotti, 2015), and also the general inclination of the civil service especially in Europe, which is an avoidance towards potential overreaction and a reluctance to listen to warnings about new threats that will require new actions.

The "infodemic" (a portmanteau of "information" and "pandemic", WHO, 2020) surrounding COVID-19 is different: its global scale is affecting numerous audiences across the planet, in dozens of languages, concurrently. As such, it is exposing the above-mentioned pattern of disinformation campaigns even to those who have ignored previous signs. If there were one potential advantage of the COVID-related crisis, it would be that the infodemic may finally open the eyes of those who may have been keeping them closed despite the warnings of the past years. Even one of the top European politicians, President of the European Commission Ursula von der Leyen, has warned of disinformation's potential for harm. As she has stated, "disinformation can cost lives" (European Commission, 2020a). This acknowledgement, unfortunately, does not mean that democratic societies are necessarily on the right track when countering disinformation. Von der Leyen's statement shows that, at present, we manage only to warn the potential victims of information aggression, but we are not doing anything about the information aggressors themselves. Her statement failed to mention the identities of the actors spreading disinformation, which are primarily the regimes of Moscow and Beijing. As a result, these information aggressors do not have to pay even the smallest price for spreading lies and threatening peoples' lives, i.e., being publicly exposed or being named and shamed for their information aggression.

According to a 2021 survey by the STEM Institute of Empirical Research, 40 per cent of Czechs believe in a few or many conspiracies about COVID-19 (STEM, 2021). For example, "more than a third of Czechs believe that the Czech government is misusing the pandemic to control citizens, that it is hiding side effects of vaccination or that hospitals are exaggerating the risks to earn money". Almost 80 per cent of Czechs, i.e., four out of five people, believe at least one of the COVID-19 conspiracies. A similar poll from autumn 2020 indicates that, in some of the surveyed countries, such as Nigeria, Greece, South Africa, Poland and Mexico, 40 to 60 per cent of the population buys into various disinformation and conspiracy narratives about the coronavirus (Henley & McIntyre, 2020). This skepticism is not just the result of one year of the COVID-19 infodemic. As previously mentioned, disinformation targeting vaccinations proliferated prior to the pandemic, and the disinformation campaigns related to COVID-19 are simply building on these previous campaigns. The Kremlin's disinformation campaigns have returned to Cold War-esque aggression following the events on the Euromaidan revolution and the Russian Anschluss of Crimea and invasion of Eastern Ukraine. The aim of these campaigns is "to undermine liberal democracy, and to sow and amplify mistrust in credible sources of information (be it governments or mainstream media), in the geopolitical direction of a country and in the work of intergovernmental organizations" (EUvsDisinfo, 2018a), and they have been systematically working towards these aims since 2014. Once these goals have been achieved, even partly, it is much easier for these actors to spread any other conspiracy, whether it be about governments falsifying elections, Russian secret service crimes and terrorist attacks abroad,[1] or COVID-19. Spreading conspiracies about one topic facilitates the acceptance of conspiracies about other topics (Kalenský, 2020a).

It is not just one campaign, one narrative or one channel that matters. Disinformation operates as "drops of water falling on a stone", in that "over five minutes, ten minutes, fifteen minutes, one hour, one day nothing happens, but five years, ten years, fifteen years, goes by, and you've got a hole in the stone" (EUDisinfoLab,

[1] As a 2018 polling in the UK showed, only one half of British population trusted the facts as presented by the government, meaning that the rest of the population was more or less vulnerable to the disinformation narratives spread by the Kremlin's disinformation machine and its witting or unwitting helpers (McTague, 2019).

N/A; Schoen & Lamb, 2012). Therefore, when analyzing the successes of disinformation around COVID-19, we cannot focus solely on the period which starts from the onset of the pandemic in 2019. The seeds were planted long before, and the soil was fertilized by previous campaigns. These false narratives have been spread in an environment that has already been primed for disinformation, and by people and institutions who already know their audiences and exploit their respective trigger points. Focusing only on the last several months is analogous to watching the final three minutes of a basketball game where one team is 30 points ahead—all of the decisive action has already taken place much before this point in time. The disinformation campaign surrounding COVID-19, as well as its apparent success, clearly indicates that our reaction towards the threat of disinformation thus far is not effective. Despite what we, in the democratic world, have done so far—the numerous reports on disinformation, as well as the pressure on social media to act—nothing has stopped information aggressors from spreading lies and from constant and systematic efforts at broadening their audience. This all points to a statement that might sound harsh, but is, in all likelihood, based on reality: we are still on a losing trajectory and, unless we dramatically change the attitude and mindset that has so far prevented us from targeting information aggressors, the problems associated with disinformation "are likely to get worse, not better," as Christopher Steele, the ex-head of MI6's Russia desk said in 2020 (Cadwalladr, 2020).

7.2 What to Avoid When Countering Disinformation

There are several potential weaknesses when it comes to countering disinformation. Most of these weaknesses, ironically, are related to some of the inherent strengths characteristic of democratic societies—like the division of power, or the tendency not to abuse state power—and might sometimes lead towards a reluctance to act. It might be useful to take note of these weakness as lessons learned to strengthen subsequent efforts.

First, we are dealing with a problem that spans across multiple traditional domains. This problem is hybrid not only in that disinformation campaigns lie somewhere in the grey area between war and peace, but also in that it pertains equally to foreign policy as it does to internal security, external security, and digital space. However, our institutions frequently do not reflect this new reality. As such, we need to avoid the mindset that puts new threats, like disinformation and other hybrid threats, into just one box, and makes such threats the responsibility of just one ministry or institution.[2]

Second, the tendency to look for silver bullet, or catch-all, solutions might detract from more targeted, effective measures that solve smaller parts of the problem. There are no magic solutions. Information aggressors use many different weapons, channels, rhetoric, and approaches for different audiences (EUvsDisinfo, 2018a). Aggressors have adopted this strategy because they know that each given channel

[2] Cf. the cross-governmental Counter-disinformation Unit in the UK (UK Government, 2021).

and approach will reach and resonate with only a segment of the population, and that other parts will need something different. We need to copy this approach when countering disinformation. We need to adopt a mindset that accepts that many countermeasures can solve no more than, for example, one to five per cent of a given problem, and that we need many different countermeasures applied simultaneously, ideally in a coordinated manner, in order to succeed.

Third, democratic states may often be significantly more hesitant to act and defend their interests than their autocratic opponents,[3] which is also demonstrated in their response to disinformation. The Kremlin's disinformation campaigns constantly evolve and adapt to new conditions and also, therefore, to some of our uncoordinated countermeasures (Newman, 2020; Snegovaya & Watanabe, 2021). Information aggressors will even experiment with rudimentary measures that show them the potential "dead ends", which provides them with the knowledge of how *not* to conduct their operations.[4] They collect data about the reactions of our populations to given information and material, up to the point where they may even know our audiences better than we know them ourselves (Kalenský, 2019a). By contrast, the countermeasures employed by Western societies are significantly more conservative and risk-averse. Far too often, we go for the easiest, least controversial measures that have the highest chance of being accepted without issue, but also tend to have the lowest chance of actually solving the problem at hand. Countermeasures like "media literacy programs" or "positive narratives" have been frequently utilized in policy debates, even dating back to 2014–2015. Despite all this time spent, solutions either have yet to show very persuasive results or, at best, they have shown that these measures are not enough.

There are many other potential weaknesses to consider, such as the constant underestimation of the threat, even in countries that are seemingly quite vigilant

[3] This disbalance between the dictator's will to act and the democrat's unwillingness to stop him gets frequently described in Garry Kasparov's overview of the Western response to Vladimir Putin's dictatorship. "We must decide what we value and decide what is worth fighting for and then – the most important part – we must fight for it. If we fail to do this we will lose to those who believe in other things, in worse things. We will lose to those who don't believe in the value of human life or liberty, and who are willing to fight to impose their dark vision of humanity on others" (Kasparov, 2015).

A similar observation can be seen in Richard Stengel's Information Wars where he quotes his interview with Geoff Pyatt, the US ambassador to Ukraine. "Right now […] we are being out-messaged by the Russians. The Russians don't have a clearance process. They don't feel the need to be truthful. We are too timid and reactive" (Stengel, 2019) (page 87).

[4] As e.g. in the Nordic countries, where they tried to open the local services of Sputnik, but they had to shut them down after a while, most probably because they had insignificant audience (Pettersen, 2016).

But that does not mean that the pro-Kremlin disinformation campaign would leave this region altogether, it only means they focus on other vectors of information aggression, like manipulating the online debates under the articles of established media outlets, cf. Aro (2015b) and EUvsDisinfo (2016).

(Bender, 2019), and the amount of both witting and unwitting helpers of disinformation campaigns (Weiss & Goldsmith, 2021).[5] However, the three aforementioned reasons seem to be at the root of many other problems and, as such, they appear to be the main obstacle in implementing a solid, robust and effective system of countermeasures.

7.3 Various Approaches to Countermeasures

Once we overcome these potential weaknesses when countering disinformation, accept that we need to act on many different levels experiment with measures and adapt accordingly, a whole universe of possibilities emerges. There is already a solid toolbox of measures that have been tried by various countries in the Euro-Atlantic space, and some recommendations for new measures have yet to implemented. Some researchers categorize these measures based on the actor that will be implementing them—whether it be government, civil society, or private sector (Fried & Polyakova, 2018). Sometimes, suggested countermeasures may be undertaken by one specific actor, whether it be government or civil service communicators (MSB, 2019; UK Government, 2019). Alternatively, some researchers divide countermeasures into groups based on their specific target audiences (Lucas & Pomeranzev, 2016). This section will categorize countermeasures according to the "Four Lines of Defense" division outlined in 2019 (Kalenský, 2019a), which is based on the desired outcome of the countermeasure, rather than the organization or actor intended to implement the measure. This approach will take into account the wide disparity between countries in the Euro-Atlantic space, in terms of their information and legal environment, and the level of awareness and experience that they have with the threat of organized disinformation.

For example, while in some countries it is the government that effectively leads counter-disinformation efforts, in other countries the government seems instead to underestimate the problem, or even worsen it outright. In cases where the government does not adequately do the job, civil society and private sectors need to step in. If they do not, information aggressors will have free reign without effective opposition.

7.4 "It's Not Just About Russia"

Before diving into specific countermeasures, it is useful to clarify one point. Many of the countermeasures mentioned below are based on the author's previous work regarding pro-Kremlin disinformation campaigns. While this approach is specific

[5] Weiss, M., & Goldsmith, J. (2021). Syria Chemical-Attack Deniers Admit Links to WikiLeaks and Russia. *thedailybeast.com.* https://web.archive.org/web/20210509092956/https://www.thedailybeast.com/syria-chemical-attack-deniers-admit-links-to-wikileaks-and-russia.

to Kremlin-based disinformation, it is applicable to other actors as Kremlin-based disinformation often serves as a blueprint for other disinformation campaigns and is often amplified by other actors. First, the Kremlin's disinformation ecosystem is far bigger than commonly perceived. Russia is responsible for roughly two thirds of foreign influence efforts documented globally (Martin et al., 2020). In other words, Russia conducts twice as many information operations against other countries as the rest of the world combined. Moreover, the Kremlin controls an astronomical number of disinformation channels that are not accessible to the majority of other actors. This extends farther than just using "trolls" on social media, but also to thousands of information professionals in the Russian government, including within its embassies all over the world and its very active secret services. Thousands of other professionals affiliated with the state's "pseudomedia" are also exporting the Kremlin's disinformation campaigns outside of Russia, including through both witting and unwitting local helpers, like certain NGOs and GONGOs (government-organized non-governmental organizations), extremist politicians (Wesslau, 2016), etc. (Kalenský, 2019a).

Russian disinformation campaigns are also increasingly focused on "information laundering", which blurs the original source of given information, and instead finds local sources that spread the disinformation for them.[6] Many seemingly "domestic" sources, an attribution usually based on nationality alone and which disregards the origin of the disinformation they are spreading, are in fact just multipliers of the Kremlin's disinformation campaigns, knowingly or not. Such sources may even be highly respected (Kalenský, 2020b; Weiss & Goldsmith, 2021), making the impact all the more powerful. As was recently made public by the US intelligence community, the Kremlin deliberately targets domestic actors to support their influence and disinformation operations (Barnes, 2021).

This mechanism of information laundering has also been present during the COVID-19 infodemic, wherein Chinese official sources have amplified disinformation narratives originating in Russia, giving them a boost (Semantic Visions, 2020). The possibilities available to the Kremlin, an actor that has launched "the most amazing information warfare blitzkrieg we have ever seen in the history of information warfare" in 2014 (Vandiver, 2014), and who continues to improve its capabilities regarding their information campaign, as well as the resultant threat of these disinformation campaigns, are simply incomparable to the mythical "400-pound person sitting on his bed" (Calamur, 2018). Moreover, other actors, in particular, China, are replicating Kremlin tactics to the degree that researchers now discuss the "Russianization" of Chinese information operations (Vilmer & Charon, 2020). There is an ever-increasing level of cooperation between Russian and Chinese state pseudomedia (European Values Center, 2021). The Kremlin is by far the most aggressive, experienced, innovative and emulated disinformation actor in the world right now. The ease and success with which they conduct their campaigns clearly inspires other such adversaries to use similar weapons, even if just on a more limited scale.

[6] A tactic that seems to be more and more important for the Kremlin each year: Kalenský (2019b), Grossman (2020), Snegovaya and Watanabe (2021), Applebaum (2021).

Second, many countermeasures will work irrespective of the specific information aggressor or disinformer at which they are directed. If we proceed to target measures towards the biggest information aggressors in the world, it will be relatively easy to adapt them when addressing smaller, less aggressive troublemakers—who might even have already been deterred by robust action taken against the bigger information aggressors.

Third, stopping the biggest information aggressor in the world, i.e., the Kremlin, would be a very strong message for other potential aggressors. If there are proven consequences for even the biggest, most experienced criminal, less experienced criminals will think twice before committing the crime.

7.5 Four Lines of Defense

Currently, there appear to be four basic "lines of defense" when countering hostile information operations.

1. Documenting the threat, in order to collect more data on the threat and gain a stronger understanding of the threat in the information environment.
2. Raising awareness about the threat to broaden the audiences that are aware of it and, thus, are inoculated against it. In contrast to the first line, which seeks to obtain more information about the threat, the second line attempts to ensure that more people have at least basic information about the threat.
3. Repairing, mitigating, and preventing weaknesses found within the information system, in order to decrease the ease of exploitation for attackers and to make their target smaller and harder to hit. This line of defense should mitigate effect of the threat on the target.
4. Limiting, challenging, constraining, punishing and deterring information aggressors. This line of defense, unlike those prior, is not directed at the victims of information aggression, but aims to decrease the aggressors' desire to be aggressive instead.

Within each of these lines of defense contain various tactical countermeasures that can be undertaken. To underline the caveat above, none of these measures will solve one hundred per cent of the problem. However, this does not mean that the given measure is necessarily useless—it merely means that we need to be applying more such measures, which constantly adapt and evolve with the changing information environment and aggressors. It is critical to act and adapt to disinformation, providing space to develop lessons learned from efforts to counter disinformation versus searching for ideal counter measures which can be unattainable. The subsequent section will discuss the tactical countermeasures relevant to each line of defense one by one.

7.5.1 Line 1: Documenting the Threat

Documentation is the imperative first step without which almost nothing else is possible; and at the same time, still a measure that is gravely underestimated. It is necessary to document: the channels involved in spreading disinformation, the disinformation narratives themselves and how they change with time, how such information travels from one medium (e.g., language, platform, "filter bubble", etc.) to another, the actors involved in proliferating them (both knowingly or not), and the effect that such disinformation has on given audiences. While this task may sound relatively straightforward, unfortunately, to date, this task has not, and is currently not, being dealt with properly. There are still many questions that we do not have answers to, which means that we are fighting in a fog, not knowing what we are standing up against.

The pioneer of systematic, day-to-day documentation of disinformation in the twenty-first century is the Ukrainian project StopFake.[7] This project started right after the Anschluss of Crimea by the Russian Armed Forces, and centers on exposing and refuting Russian disinformation campaigns abroad. The project's focus is specifically on disinformation campaigns targeting Ukraine, but StopFake has also debunked disinformation surrounding other countries or individuals.

The process of debunking may sometimes be quickly dismissed as too laborious, or as a "whack-a-mole" approach. There is also an incorrect, though unfortunately relatively popular, myth that debunking strengthens, rather than weakens, disinformation efforts (Global Engagement Center, 2020c; Swire-Thompson et al., 2020). However, in order to be able to label a piece of information as disinformation, it is necessary to fact-check the given story and to identify whether and where, if at all, the piece distorts reality, i.e., refute the disinformation piece, debunk it. There is no other way to identify disinformation. Without the process of debunking, it is not even possible to have a conversation about disinformation, let alone to appropriately document instances of disinformation. Without the documentation of cases of disinformation, there are no facts on which a decision or future learning may be based.

Documenting disinformation, no matter how underestimated or dismissed, is the imperative first step towards countering disinformation, without which nothing else could be done in an informed and verifiable manner. It is no coincidence that much of the misunderstanding and underestimation of disinformation campaigns and how to counter them comes from individuals and organizations who have never undergone this critical first step. It is also not unusual that those experienced with the vast disinformation ecosystem often have to point out that their respective societies continue to underestimate the scope of Kremlin disinformation and propaganda.[8]

[7] https://www.stopfake.org/en/main/.

[8] Bender (2019), or Owen (2018). The author of this piece has heard many times since 2015 from various European civil servants that at first, their impression was that those who are warning against the threat of disinformation are just exaggerating; but once they started monitoring the

StopFake's approach to countering disinformation has been emulated around the world. After the Kremlin's Anschluss of Crimea and the renewed Russian information aggression that followed, NATO has created a factsheet called "Setting the Record Straight" (NATO, 2020), which debunks some of the most notorious disinformation narratives spread by the Kremlin and its disinformation ecosystem. Since 2015, the EU's counter-disinformation unit, the East StratCom Task Force (EEAS, 2021), has been publishing its "Disinformation Review", a project that documents pro-Kremlin disinformation,[9] and has since evolved into the EUvsDisinfo campaign. As of April 2021, the Task Force collected over 11,000 cases of disinformation, which are presented within a searchable database[10]—so far, the only such database of its kind in the world.

However, the work of EUvsDisinfo also indirectly shows us how much we are missing. As of spring 2021, Brussels has not replicated this model beyond addressing pro-Kremlin disinformation campaigns. There is no such project, for example, that documents disinformation originating in China.[11] Rather, the coverage of Chinese Communist Party disseminated disinformation seems to be based on the work of non-EU researchers and journalists, as opposed to the EU's own data set (EUvsDisinfo, 2020b). Similarly, hard data about disinformation campaigns such as that of ISIS, or other so-called domestic actors, are not provided by the EU.

The refuting and documenting of disinformation is also done by think tanks, NGOs and civic activists. The Atlantic Council's Digital Forensic Research Lab focuses on "identifying, exposing and explaining disinformation using open source research"[12] and in 2021, the team presented a report on COVID-related disinformation (DFRLab, 2021). The Prague-based European Values Center for Security Policy (European Values) exposes "Russian influence and disinformation operations focused against Western democracies."[13] The Lithuanian "Elves"[14] kick-started their counter-trolling efforts following the Russian invasion of Ukraine. Through their cooperation with established media and information technology and artificial intelligence (AI) experts, a unique system dubbed Debunk.eu has been built (Global Engagement Center, 2020a). Debunk.eu combines an algorithm which identifies potential disinformation stories with robust manpower to verify the information pinpointed by the algorithm, i.e., fact-checking and setting the record straight (Delfi, 2020). The Czech company Semantic Visions also uses AI to monitor the spread of disinformation stories (Semantic Visions, 2020).

disinformation campaigns in their country, only then they realized that it is in fact much bigger than they originally had thought.

[9] https://euvsdisinfo.eu/disinfo-review/. Full disclosure: the author of this piece is also the founder of the Disinformation Review and the EUvsDisinfo campaign: https://euvsdisinfo.eu/.

[10] https://euvsdisinfo.eu/disinformation-cases/.

[11] Despite the calls from the European Parliament for such a team: Fotyga (2020).

[12] https://www.digitalsherlocks.org/about. Full disclosure: the Atlantic Council's DFRLab is the author's current place of work.

[13] https://www.kremlinwatch.eu/#about-us.

[14] Meaning those who are fighting the online „trolls".

The Czech version of the “Elves”[15] are documenting what is happening across disinformation-oriented outlets, and also across non-public channels that spread disinformation, including chain emails (which are still popular among the elderly population in the region; EUvsDisinfo, 2017a). Researchers in Ukraine have documented the rise of pro-Kremlin disinformation that has penetrated some of the country’s most popular Telegram channels (Velychko, 2020). Infiltrating such groups that use closed channels of information is probably best done by actors within civil society, rather than by formal institutions (governments and other such organizations would be susceptible to accusations of “Big Brother”-like espionage). Apart from the work undergone by the public sphere, there are also institution- and government-level efforts underway (Pogrund, 2020). This work, though sometimes very high in quality, may be limited in impact if the work remains non-public.

Even though there are already a few initiatives working on documenting and refuting disinformation, there are still many crucial questions that remain unanswered. For example:

- How many disinformation channels are there? Is their number increasing or decreasing? How many new channels have appeared in the last week/month/year?
- How many messages do these channels spread on a daily/weekly/monthly basis?
- How many people do these messages target and reach?[16]
- How many people believe the messages of each given disinformation campaign? (because, hypothetically, the decreasing numbers of the disinformation narratives or channels does not necessarily mean that the campaign would be less successful, it can just mean it is evolving into a more effective stage, doing the same job with less resources)

Without answers to these questions, we remain fighting in a fog. We are in a situation akin to a country that has been attacked by an outside aggressor but does not have even the remotest of ideas about how many soldiers, tanks, jets or missiles the enemy has in their arsenal, or even how many people per day they succeed in killing.

In addition to this, without having the answers to these questions, we cannot address other crucial questions. Following every major incident that the Kremlin has lied about, many researchers and civil servants receive the same question: Has there been an increase in disinformation? The issue here is that, if we are not properly measuring the extent to which disinformation is present during the uneventful, or “calm”, periods, we cannot begin to address a potential increase, or lack thereof,

[15] https://cesti-elfove.cz/.

[16] One element of the weak reaction of the Western democracies to the threat of disinformation is the focusing on the singular elements of the disinformation campaign while not seeing the whole of it. In other words, not seeing the wood for the trees. It is frequently possible to hear the dismissive “But this article / social media post / pro-Kremlin mouthpiece couldn’t have made the difference.” It is not one measure that matters, it is the sum of them. If we had information like “every week, there are 20,000 disinformation channels spreading 200,000 disinformation messages in 30 different languages targeting 3 billion potential readers and viewers”, it would be information impossible to ignore even by those who are skeptical about or ignorant of the magnitude of the threat. But, because of the lack of documenting of the threat, we do not have it.

following major events, or “hot” periods. We do not know what the baseline is and, therefore, we cannot talk about deviations from the baseline, empirically speaking at least. We cannot answer questions about the potential success, or lack thereof, of various disinformation campaigns, either, if we do not have empirically certain data about them. Essentially, there is an unknown, but powerful, force influencing our audiences, and we cannot say how big or how dangerous it actually is.

We are in a situation in which we see but the proverbial tip of the iceberg, without having a clue as to whether we are aware of ten or only one per cent of the problem. We need to be designing various solutions to address one hundred per cent of the problem. Finding any missing information as soon as possible should be the imperative. Governments should be taking the lead or, at the very least, be providing the necessary funding for relevant third-party actors, as many civil society organizations do not have sufficient resources for adequate monitoring.

Aside from trying to document disinformation campaigns as exhaustively as possible, we also need to document the success, or lack thereof, of each given campaign, while answering the following question: “How many people believe each disinformation message?” Unfortunately, the work that is being done to address this question is quite sporadic and less systematic than required.

The question surrounding the impact of disinformation campaigns is one of the most difficult in the field. Disinformation expert Ben Nimmo, the head of Global Threat Intelligence Strategy at Facebook, has developed a model based on whether information remains on one platform, or travels across multiple, and whether this information similarly remains in one community, or spreads across many (Nimmo, 2020). Other researchers have measured the engagement rate of specific disinformation narratives on social media, often concluding that disinformation frequently beats true information in this respect (Dizikes, 2018; RFE/RL, 2020a).

Another option is to measure how many people believe a particular disinformation campaign via opinion polls—because it is highly probable that the information aggressors are more interested in turning peoples’ hearts and minds, rather than in engagement numbers on social media. This method is more precise in the cases where we know that a particular message would never be in the information space if it had not been introduced by various information aggressors, such as the pro-Kremlin disinformation campaign surrounding the crash of flight MH17, or the poisonings of Litvinenko, Skripal and Navalny. In instances where a particular disinformation campaign builds upon pre-existing conditions, such as low trust in authorities or fear of various socio-economic groups, it is hard to distinguish between the impact, or lack thereof, of the campaign itself, as opposed to other factors. That being said, this can still provide valuable data to consider for policy makers or opinion leaders dealing with the information space.[17]

[17] Cf. data from a Czech opinion poll in 2019, showing e.g. that 70 per cent of population believes the false narrative that the EU dictates Czech Republic what to do and violates its sovereignty (at the same time, around one third of respondents believe that Russia and China deserve respect even if they interfere into Czech affairs), or 50 per cent of people believing that the EU is organizing the migration wave: Spurný (2019).

A cause for concern is the tendency for significant numbers of the population in several European countries to fall for disinformation narratives (Shutov, 2020). For example, even in a country like Ukraine, where the level of awareness about the Kremlin's disinformation campaigns is very high, there remains a significant part of the population that believes disinformation narratives that would not appear within the information space if not for the pro-Kremlin disinformation ecosystem (EUvsDisinfo, 2017b). Unfortunately, at present, there are no individual actors or organizations, public or private that are conducting regular opinion polls on the subject.[18]

As of yet, we do not have consistent, reliable, and replicable information about the potential successes of these disinformation campaigns. Not only does this mean that we are uncertain as to how big the threat is that we face, we also cannot reliably measure whether our countermeasures work (i.e., whether the number of people believing various disinformation narratives have decreased).

Documenting the threat of disinformation may seem like a measure that is critical only after a given event or campaign has occurred but, in fact, this measure can be of use even during an ongoing disinformation campaign. Following the poisonings of Sergei and Yulia Skripal by Russian secret service agents, the British government documented and exposed the various conflicting, constantly changing narratives spread by the Kremlin as they appeared (Dettmer, 2018). In so doing, the British government was swift and consistent in its response to the disinformation, and the conflicting messages and lies spread by the pro-Kremlin disinformation ecosystem were exposed (EUvsDisinfo, 2018b). A simple timeline of the various messages spread by Kremlin's propaganda bullhorns made it obvious as to the deception at play.[19]

7.5.2 Line 2: Raising Awareness About the Threat

Documenting the threat is the first step to countering systematic disinformation campaigns, and the raising of awareness afterwards, by making use of the data collected, is the natural continuation of this first step. Projects such as the DFRLab (Atlantic Council), EUvsDisinfo (EU), and Kremlin Watch (European Values) are not only documenting cases of disinformation, but are also publishing relevant articles and reports, as well as providing training and briefings for governments, media, and the academic community. However, there is a limit to these organizations' reach. Irrespective of whether the actor is governmental or non-governmental, neither can reach nor persuade an audience in full—they each have a limited target audience.

[18] Eurobarometer would be the natural choice in Europe, but they do not seem to measure how many people believe in this or that disinformation message: European Commission (2020b).

[19] A simple timeline of the messages of the disinformation ecosystem is a tool that works particularly well in cases where the information aggressors try to bury the truth under a pile of contradictory statements, see e.g. EUvsDisinfo (2018b, 2020c, 2020d).

It is necessary not only to collect more data about disinformation campaigns and to educate one's own audience about the threat, but also to constantly broaden the audience that receives that message.

In addition to the organizations mentioned above, the Swedish Civil Contingencies Agency (MSB) created a handbook for civil service communicators intended to help them counter information influence activities (MSB, 2019). A similar effort was initiated by the UK government (UK Government, 2019). The MSB and the Security Service Säpo have also been briefing politicians and media in Sweden (EUvsDisinfo, 2018c). The Czech Center for Countering Terrorism and Hybrid Threats is also focusing on educating the country's civil service about hybrid threats, including disinformation (Ministerstvo vnitra ČR, 2016). In some countries, it is the secret services that provide excellent information on the threat, and their materials are highly valued by the expert community and the media—see e.g., the annual reports of the Estonian KAPO,[20] or the Lithuanian National Threat Assessment.[21]

The NATO StratCom Centre of Excellence (CoE) in Riga publishes reports on the topic of disinformation and strategic communication.[22] They also organize conferences targeted at civil servants and experts dealing with the topic. The CoE also provides briefings and training for governments in NATO's member countries. The joint EU-NATO Hybrid CoE in Helsinki is working in a similar manner.[23]

Ideally, increasing awareness on a given disinformation campaign or narrative may help to mitigate or dispel a false story before it has time to spread. A comparison between the infamous "Lisa case" in Germany with its Lithuanian counterpart demonstrates the difference between how a disinformation campaign plays out in a country aware of the threat, versus in a country where this awareness is not very high. The Lisa case took place in Germany in January 2016 (Meister, 2016). A 13-year-old girl from a Russian-speaking family, upon not making it home in time for her curfew, told her parents that she had been raped as an excuse. While she later admitted that the story was false, and the German authorities confirmed this, pro-Kremlin media and activists in Germany ignored this and continued to spread only the original falsehood. This resulted in demonstrations and a public denunciation by Russian Foreign Minister Sergei Lavrov of the German authorities, accusing them of a cover-up (McGuiness, 2016).

By contrast, a year later, a similar fake story in which an underage girl was allegedly raped appeared in Lithuania, failed completely (Schultz, 2017). This was in large part due to the Lithuanian Armed Forces' StratCom team's preparation of the country's civil servants and local authorities for such attacks, which had been anticipated prior to the NATO Enhanced Forward Presence exercise in the country. Local authorities, once having registered the falsity of the planted story, alerted the StratCom team, which quickly debunked it and alerted all relevant stakeholders. The first Lithuanian media coverage on the subject was not "a little girl allegedly

[20] https://www.kapo.ee/en/content/annual-reviews.html.

[21] https://www.vsd.lt/en/threats/threats-national-security-lithuania/.

[22] https://stratcomcoe.org/publications.

[23] https://www.hybridcoe.fi/.

raped", as with the German counterpart, but rather "another information attack against Lithuania". An extremely high awareness of disinformation campaigns in the country, as well as a timely exposure of the false story (which by definition had to involve the often-dismissed debunking of the story), resulted in one of the most instructive success stories in the entirety of the EU thus far.

However, given the fact that audiences today are more fragmented than ever before, it is necessary to move beyond the small circle of government officials and experts focused on the topic, and try to raise the awareness among broader audiences as well. One proven way to attract wider audiences is through humor. In fact, the ridiculous nature of some of the falsehoods spread by disinformers may provide a lot of ammunition for jokes and ridicule. The tweet by the Canadian Permanent Representation to NATO, which ridiculed Russia's repeated "mistakes" regarding geography, has already become well-known (@CanadaNATO, 2014; Burtch, 2018). The official account of the Ukrainian presidential office mocked the Kremlin's propagandistic appropriation of Ukraine's history as being a part of Russian history with a GIF from The Simpsons (Thomsen, 2017). The EUvsDisinfo project has a rubric for lighter, sarcastic pieces that make fun of recent propaganda.[24] Taiwan's Prime Minister has also used funny memes directed towards COVID-related disinformation (Yang, 2020).

Interestingly, research shows that act of ridiculing can work well with general audiences, although it is not advised when appealing to conspiracy theorists (Lewandowsky & Cook, 2020). In spring 2021, EUvsDisinfo published a specialized guide to communicating with those who believe the COVID-related conspiracy theories (EUvsDisinfo, 2021a). The US State Department's Global Engagement Center (GEC) details the use of a "counter-brand approach", in which those countering disinformation should not refute every single false claim, but rather point out the untrustworthiness of the source spreading disinformation, thus reaching broader, mainstream audiences (GEC, 2020b).

Notably, professional influencers are likely the most effective at broadening their audiences and reaching beyond the typical filter bubbles of civil servants and NGO activists. One such success story once again takes place in Lithuania; a local TV journalist, Andrius Tapinas, launched a show named "Hang in There" (Schearf, 2017), in reference to the ridiculed response of then Russian Prime Minister Dmitry Medvedev to a Crimean pensioner who confronted him about the low pensions in Russia (RFE/RL, 2016). As Tapinas himself describes, the aim of the show is to make fun of the Russian government and politics. In the Czech Republic, the Youtuber known as Kovy has filmed a series of videos aimed at increasing media literacy among the country's younger generation, as well warning about the dangers of manipulated information (Jeden svět na školách, 2019). Kovy's primary audience are Czech youth—an audience primarily untouched by both government, NGO and CSO initiatives on the project.

It must be acknowledged that some methods of communication may work effectively for only a part of a given audience and may be ineffective for another—but this

[24] https://euvsdisinfo.eu/category/fun/.

is not a justification to stop implementing these various methods. As journalist Anne Applebaum documented prior to the onset of COVID-19 in Italy regarding the anti-vaccination disinformation campaign, different approaches to raising awareness of a threat will reach and impact different people, but this does not make one approach better or worse than another (Applebaum, 2019). Efforts towards finding new opinion leaders to reach new audiences should never stop—there is always a means to make a given message resonate.

7.5.3 Line 3: Preventing and Repairing the Weaknesses in the Information System

This line may appear to resemble what Russian disinformers regard as "defensive measures" (Giles, 2016), but democratic societies must avoid the Kremlin's approach in this respect. Putin's regime seeks to supposedly "defend" the information space via measures that, for example, make it illegal to talk about historical facts that contradict Kremlin-propagated narratives (Meduza, 2016), restrict independent media in Russia, and attempt to cut its population off from independent sources of information. This, of course, should not be pursued by the West. It is, however, useful to acknowledge that some of "weak" areas within the information space are systematically abused by information aggressors.[25]

Line 3 slightly overlaps with Line 2, in that raising awareness effectively can repair some of the weaknesses found and exploited within the information system. However, some of these weaknesses require more than just increased awareness.

The ways in which Kremlin propaganda has abused the existence of racial tensions in the US and around the world is well documented—be it during the Cold War (McCauley, 2016; Rid, 2020), or during the recent US Presidential elections both in 2016 and 2020 (Green-Riley & Stewart, 2020). Similarly, Kremlin disinformers attempt to identify and aggravate tensions across any given target audience—including within the context of the LGBTQ+ community, migration issues, global warming, generational gaps, urban and rural divides, or historical and ethnic tensions across and within countries (EUvsDisinfo, 2018a). Political divides are an extremely fertile soil for disinformation (Bugajski, 2020). Any topic that has the potential to divide or polarize target audiences, as well as to make the debate emotional instead of rational, works, "because an audience shaken by strong emotions will behave more irrationally and will be easier to manipulate" (EUvsDisinfo, 2018a). The disinformation related to COVID-19 is also exploited for the same reasons (Kalenský, 2020a). The work to build bridges between various alienated groups and to reduce tensions, instead of aggravating them, requires much more than just efforts at communication, and has to be done by relevant topical policy makers. However, it is useful to keep in

[25] "Anti-democratic forces use misinformation, disinformation, and weaponized corruption to exploit perceived weaknesses and sow division within and among free nations, erode existing international rules, and promote alternative models of authoritarian governance" Biden (2021).

mind that such work has the potential to reduce targets for information operations; in other words a less polarized, less hysterical and more rational society means that the information aggressors do not find such a fertile soil for their operations.[26]

Another part of the information space that is especially susceptible to abuse is reliable traditional media. The New York Times, in reference to the 2016 Democratic National Convention email leak, acknowledged that it, as well as other media outlets, have become a tool of Russian intelligence (although the New York Times is one of the few media outlets to have admitted this mistake; Lipton et al., 2016).

Unfortunately, this accountability and awareness regarding traditional media's role, even unintentionally, in the support of malign disinformation campaigns has not necessarily carried over to other such outlets. Throughout the COVID-19 crisis, European media outlets have frequently re-published statements made by Russian authorities, e.g. regarding the Sputnik V vaccine, without making note of the fact that the same individuals are responsible for spreading false information surrounding Ukraine, Syria, the downing of MH17, the crimes orchestrated by Russia's secret service, state-sponsored doping, or pretty much any scandal Russia had been involved in the past several years (France 24, 2021; Henley, 2021; Independent, 2021). Pro-Kremlin outlets have also managed to alter the information environment in Italy via reliable media outlets to the extent that even an independent court appeared to rule in favor of the information supported by the disinformation campaign, instead of with factual evidence (Tokariuk, 2021).

Presenting the statements of these representatives, especially within the context of their disinformation campaigns, without making it clear that they are extremely untrustworthy sources of information, legitimizes them and, in effect, misleads the audience to the benefit of the information aggressors.

In 2014, journalists Peter Pomerantsev and Michael Weiss proposed that a Disinformation Charter for traditional and social media be formulated in order to distinguish between acceptable and unacceptable behavior. They also recommended that media outlets hire specialized disinformation editors, so as to prevent media outlets from becoming inadvertent purveyors of disinformation (Pomerantsev & Weiss, 2014).

Unfortunately, there does not appear to be much progress on this front, at least not globally; however, there are success stories on the local level: the Czech Endowment Fund for Independent Journalism has created a rating of the credibility of the media (NFNZ, 2020), and the number one search engine in the country, Seznam.cz, is warning users when an untrustworthy outlet appears in their search results (Kapuciánová, 2020).

Similarly, social media is another component of the information space that is frequently abused for the purpose of spreading disinformation. For example, in

[26] A similar argument is made by Anstead (2021): "[F]ake news does not create divisions, but rather catalyses pre-existing political and social conflict. It is these divisions – whether social-economic, religious or racial – that provides the raw fuel for fake news. Tackling these inequalities will greatly undermine the purveyors of fake news."

relation to COVID-19, the verified Twitter account of the Sputnik V vaccine regularly spreads disinformation and groundless accusations directed towards various European institutions (EUvsDisinfo, 2021b). Unfortunately, this account is not the first instance in which a social media platform has "verified" and/or recommended content from a notorious disinformation channel (Harwell & Timberg, 2019). The problem of information manipulation occurring on new media is not solely limited to the biggest social media platforms—it also concerns other online resources, e.g. Wikipedia (Dewey, 2014; EUvsDisinfo, 2021c). Pro-Kremlin information operations have also "appeared on Instagram, Stitcher, Reddit, Google+, Tumblr, Medium, Vine, Meetup, and even Pokémon Go" (Linvill & Warren, 2019).

Social media platforms should be pushed to clearly label and warn against content from outlets notorious for spreading disinformation,[27] instead of promoting them. A similar approach has been used against tobacco, wherein toxic substances are labeled—the smoker is still permitted to smoke, but he is also clearly warned that he is using a harmful substance. The same approach is used to label unhealthy food, and it has been known to have "overwhelmingly positive" results (Michail, 2019). Similarly, disinformation outlets may be considered harmful to one's ability to perceive reality. Research shows that there is significant potential in replicating this approach (Helmus et al., 2020).

The mass-proliferation of disinformation should not be considered a problem limited to social media.[28] As highlighted many times above, it is a problem that spreads across many other channels, including traditional media, extremist politicians, seemingly independent activists, etc. Notably, even if the big platforms attempted to remove toxic content, as in the case of the "Plandemic" conspiracy video, this does not necessarily mean that this content would disappear—it may simply move elsewhere (EUvsDisinfo, 2021d; Kharazian & Knight, 2020). However, again, this does not render the pressure on social media platforms useless, but rather indicates that this is merely one of many solutions that should be implemented in due course.

The EU has worked with the social media industry through the creation of a voluntary Code of Practice on Disinformation (European Commission, 2020c). Several platforms including Google, Facebook and Twitter have agreed upon self-regulatory standards to fight online disinformation. However, the EU Commissioners were not fully satisfied with the initiative's progress (European Commission, 2019a), and the EU Commission started working towards adding regulatory and co-regulatory measures, too. Once in force, the Digital Services Act and Digital Markets Act

[27] "Freedom of expression for platform users entails more than the right to speak. It also involves the freedom to seek and receive information, as well as the freedom to form opinions. If the ability of platform users to form political opinions is distorted by rampant political disinformation, an important dimension of their free expression rights will be undermined. Commitment to users' free expression does not require that platforms let disinformation flow, but instead justifies efforts to combat it" (Donahoe, 2020).

[28] See e.g. the clear focus on the digital aspect in the plans of the EU for the next five years: EUDisinfoLab (2020).

(which are, as of May 2021, being negotiated in the Council and European Parliament) will oblige the platforms to be more transparent about how content is ranked and advertised, and increase accountability of the platforms in response to societal risks (EDRi, 2020). In addition, the Commission will issue guidance setting out goals for strengthening the above-mentioned Code of Practice and set up a permanent framework for the monitoring of the code. Prominent social media platforms, such as those mentioned above, are frequently described as one of the largest problems when it comes to the dissemination of disinformation, and, as such, they should be doing more to mitigate this issue. However, it is fair to point out that, in recent years and especially during the COVID-related infodemic, some of these platforms have stepped up their counter-disinformation activities. Some would suggest that they are probably more proactive than many governments and other actors in the information space (Byers, 2021; Culliford, 2020; France24, 2020; Hern, 2020; Kelion, 2020; Taylor, 2020).

Another way of preventing or repairing weaknesses within the information space is to build so-called “positive narratives” (EUvsDisinfo, 2017c; Laity, 2015), which defend targeted institutions and help build resilience against attacks by information aggressors. For example, the East StratCom Task Force is also involved in “promotion of EU policies towards the Eastern Neighbourhood” and supporting independent media in the region (EEAS, 2021). In a similar fashion, the Estonian government launched a TV channel in 2015 aimed at the local Russian-speaking minority to provide them with a trustworthy Russian-language source of information, as opposed to the pro-Kremlin pseudomedia (DW, 2015). Notably, some Kremlin disinformation professionals have not seemed to be persuaded that positive messaging is an effective mechanism.[29]

Another weakness that is susceptible to exploitation is the general lack of media literacy within many targeted societies. However, a success story in this regard is provided by some of the Nordic states, which are frequently cited as examples of highly media-literate societies. The four websites of Russian government-sponsored media source Sputnik in all the Nordic countries had to shut down fairly quickly because they did not attract a large enough audience (Pettersen, 2016). In particular, Finland is often cited as one of the best examples of a highly media-literate society resisting fake news (Mackintosh, 2019), with an anti-fake news program established as early as in primary school (Henley, 2020). The Estonian government requires a 35-h media manipulation class for all secondary students (Swab, 2020).

However, the positive results that may be expected by improving media literacy within a given population are often overemphasized. First, this will require highly organized campaigns across numerous educational systems that will have to endure for years to have the desired effect, and we need solutions more urgently. Second, those who believe and/or spread disinformation may not be doing so only because

[29] “RT was created in 2005 as ‘Russia Today,’ with the mission of trying to explain the country to the rest of the world, but Simonyan told me that she soon gave up on that effort. ‘Been there, done that!’ she said. ‘We’re over with that. We don’t think that works’” (Dougherty, 2015). Approximately since the 2008 war in Georgia, Russia Today focuses more on presenting negative stories about the West and attacking and denigrating its enemies: Nimmo (2018), EUvsDisinfo (2018a).

they are illiterate (Partin, 2020). Finally, the information aggressors are not static—they can adapt to their environment very skillfully. If they are unable to exploit one weakness, they may simply move on to exploit others. In the case of Nordic states, this may be via cyberattacking, the manipulation of discussion forums under articles of established media outlets (Aro, 2015a; EUvsDisinfo, 2016; Mierzyńska, 2021), or online trolling. Potentially the worst case of targeted "online harassment" in the EU was that against the Finnish journalist Jessikka Aro, who had been exposing Kremlin influence operations in Finland. This case has already led to criminal charges for the offender(s) (BBCNews, 2018).

It is crucial to constantly investigate and evaluate the techniques and methods of information aggressors to get familiar with the weaknesses they can exploit. The "Surkov leaks" investigation has shown that the number of tools and weak spots that the Kremlin was capable of using in Ukraine was very high, beyond what the non-expert audience would imagine (Shandra & Seely, 2019). Although repairing weaknesses is a very important part of the process, it is by far not enough. "[A]lmost by virtue of their open nature, democracies will have societal vulnerabilities" (Ellehuus, 2020). Every society that wants to remain open and democratic will necessarily be exposed to possible disinformation attacks. Being informationally vulnerable is an inseparable part of being democratic—which is a principle that the information aggressors understand significantly better than the democratic societies themselves.

Therefore, it is essential to limit the willingness of the aggressors to conduct these attacks.

7.5.4 Line 4: Punishing, Deterring, and Limiting the Information Aggressors

An unpunished act of aggression hits the victim at least three times. First, the victim must deal with the unpunished aggressive acts themselves. Second, each aggressor who faces no resistance is only encouraged towards further aggression—a weak or non-existent reaction only serves as confirmation that the act of aggression is the right choice, which results in likely future aggression. The Kremlin's behavior over the past few years is the best evidence for this claim. Third, inaction demonstrates to every potential future aggressor that we tolerate this behavior, increasing the risk of future aggression from other actors as well. The ways in which the Chinese regime builds on and spreads Kremlin's COVID-19 disinformation confirms this.

The need to start "imposing costs on perpetrators" has been recognized as essential, particularly by the EU which, in December 2020, introduced the European Democracy Action Plan (EDAP; European Commission, 2020d). It remains to be seen how this "imposing costs" will be implemented. The expert community, including some government institutions, have been calling for this recognition for several years (Kalenský, 2019a; Kube, 2020). Several nations have already demonstrated before EDAP's introduction that there are many possible measures that can be undertaken to limit or to penalize the aggression of disinformers.[30]

There are both legal and non-legal measures, with the latter dependent upon the decisions of individuals and institutions. The easiest of the non-legal measures is the "name and shame" strategy. For example, as of 2021, it is widely acknowledged that the Czech President Miloš Zeman frequently acts in the interests of the Kremlin, rather than in the interests of his own country. However, in 2016, this was not yet understood and, as such, there had to be a first to publicly "name and shame" Zeman—in this case, it was European Values (Janda, 2016). This process may discourage some information aggressors, although, even if it does not,[31] this method may also help to raise awareness about the given disinformation campaign and broaden the audience aware of the threat.

Both individuals and organizations have the ability to deny access to the "pseudomedia" that serve various disinformation campaigns. If an outlet or a quasi-journalist regularly spreads lies, there should be no reason to treat them as you would a reliable, trustworthy media outlet that has made the occasional mistake. This applies to both government institutions as well as private businesses, including social media platforms. As Barack Obama has said, "The First Amendment doesn't require private companies to provide a platform for any view that is out there" (Goldberg, 2020).

In 2017, Estonia decided that the Kremlin's information weapons (which claim to be "media") were not welcome at their media press conference about Estonia's upcoming EU Presidency. This reasonable decision was questioned by the Organization for Security and Co-operation in Europe (OSCE; RFE/RL, 2017), which seemed to be completely unaware of the real purpose of the Kremlin's pseudomedia, i.e., not to report facts, but to act as an "information weapon" of a regime that is hostile towards the EU, NATO and Estonia (Nimmo, 2018). Fortunately, the OSCE's decision did not influence countries such as the United Kingdom, which, in 2019, also denied access to the representatives of the Kremlin's news agencies to a conference on media freedom (BBCNews, 2019a, b).

During the pandemic, various social media platforms have started to remove content containing COVID-related disinformation. Although the implemented filters are far from perfect (Dotto & Morrish, 2021), it shows that denying disinformers access to these platforms and removing disinformation from the social media information space is a possibility that is not limited to a few of the most aware governments; and it should be performed also by private companies, including traditional media.

An example from the US shows a creative approach by a government agency for which no special law is needed. Before the 2018 US midterm elections, the US Cyber Command targeted Russian operatives in an attempt to deter them from conducting

[30] One of the best examples is once again provided by Lithuania: Keršanskas (2021).

[31] In April 2021, the Russian investigative journalist and expert on secret services Andrei Soldatov has pointed out, in an interview about the case of GRU bombings in Czech Republic, that the Kremlin does not care anymore about the damage to their reputation, since it cannot get any lower (Soldatov, 2021).

disinformation operations that may have interfered with the elections (Barnes, 2018). With a similar aim in mind, the US conducted cyber-attacks against Russia's electric power grid (Sanger & Perlroth, 2019).

Among the legal measures undertaken, most of the examples are provided by the countries on the frontline with Russia—the three Baltic republics and Ukraine. In 2016, Lithuania used a hate speech-targeted law to address a local disinformation-spreading media outlet (Wang, 2019). Lithuania has also banned one of the Kremlin's "pseudojournalists" from the country, citing it a "threat to national security" (RFE/RL, 2019), and similar measures may be seen in Latvia (UAWire, 2020) and Ukraine (UNIAN, 2020). Latvia temporarily banned a pro-Kremlin channel due to the incitement of hatred in 2019 (European Commission, 2019b). This measure was used again to address the same channel, after violations of the Latvian law related to incitement to hatred, incitement to violence and provocation of military conflict (LSM.lv, 2021).

Due to the ongoing military conflict with Russia, Ukraine has banned some of the channels abused by Kremlin propagandists, including the previously popular social network VKontakte ("Russian Facebook"; Antoniuk, 2020). The country has also banned channels controlled by a local pro-Kremlin politician, Viktor Medvedchuk (Українська правда, 2021). Bulgaria has targeted a local pro-Kremlin politician for spreading false information surrounding COVID-19, which was deemed a threat that could cause panic (RFE/RL, 2020b). In Slovakia, the editor-in-chief of a disinformation-oriented outlet had been found guilty of defamation of race, nation and religion (Spectator, 2019). Similar rhetoric is often employed by disinformation disseminators,[32] which regularly spread hatred, incite violence and spread false stories that have the potential to cause panic or chaos. These disinformers are often acting in violation to several such laws, and their actions cannot be excused by the application to freedom of speech. In 2019, Taiwan passed an anti-infiltration law, which is intended to help the country combat Chinese influence operations (Lee & Hamacher, 2019).

Apart from these overarching laws, there is also the legal instrument of sanctions. Both Estonia and Latvia have used sanctions against the chief Kremlin propagandist, Dmitry Kiselyov (EUvsDisinfo, 2019), to make the operations of the pseudomedia under Kiselyov's purview more difficult. Estonian banks and officials refuse to deal with representatives of Sputnik, and have threatened criminal prosecution, which has forced those pseudojournalists to leave the country (ERR, 2020). Latvia has also suspended one of the websites that serves the Kremlin's disinformation campaign and, despite the relatively easy registration of a new domain, the propagandists' outreach has significantly decreased (Aleksejeva, 2020). Latvia has also suspended websites connected to Kiselyov (Tanner, 2020) and another attributed to the sanctioned Russian national, Yuri Kovalchuk (LSM.lv, 2019). Sanctions are a tool that can, and should, be used much more often when it comes to tackling disinformation.

[32] https://euvsdisinfo.eu/disinformation-cases/.

So far, the EU has managed to sanction only one pseudojournalist, Dmitry Kiselyov.[33] By contrast, his colleague, Vladimir Solovyov, who regularly spreads disinformation against the West—including COVID disinformation (EUvsDisinfo 2021e)—on the same channel and during the same evening as Kiselyov, has an Italian residence permit and owns two villas at Lago di Como (Belsat, 2019). People who lied about the Kremlin's Anschluss of Crimea and thus helped to weaken the Western response received medals from Vladimir Putin for "objective" coverage on the Crimea events, i.e. for lying and misleading the international community (Bigg, 2014). "Journalists" awarded by the Kremlin regime[34] spread harmful lies to incite hatred towards the West and its representatives Westerners on a daily basis. There is no reason to allow such people to enjoy the perks of life in freedom and democracy.

Sanctions do not need to be limited to the individual; they can be targeted against disinformation organizations as well. Many disinformation-oriented outlets have nothing to do with the work of real media, but instead are a form of information weapons– leading Kremlin's pseudojournalists like Margarita Simonyan or Dmitry Kiselyov are proud about it and describe their jobs as a cheaper alternative to killing people (Kalenský, 2019a).

Sanctioning the disinformation-oriented channels would mean they would lose money from advertisements from Western companies. Currently, the biggest advertisers on Russian state TV are Western companies like Pepsi, Nestle or Reckitt (Sostav, 2020). Because Russian state TV is the building block of the Kremlin's whole disinformation ecosystem, Western companies are effectively paying for the creation and spread of anti-Western disinformation. There is also a lot of advertisement-related money within the disinformation ecosystem, not just on Russian TV (GDI, 2021). If governments continue not to sanction disinformation-oriented channels by law, it may become the prerogative of civil society to take matters in their own hands. In the Czech Republic, local activists and PR professionals try to pressure companies not to advertise on outlets known to spread disinformation (it is worth reminding that, in order to know about which outlets these are, it is necessary to identify and document disinformation cases, in accordance with the efforts described in Line 1; Brokes, 2020).

Limiting advertisement money spent on outlets known for spreading COVID-related disinformation has also been mentioned in an EU communication called "Tackling COVID-19 disinformation—Getting the facts right" (Kalenský, 2020c)—it remains to be seen whether this measure will be implemented. In April 2021, the US decided to sanction several minor disinformation-oriented outlets operated by Russian secret services (U.S. Department of Treasury, 2021).

[33] The only other „media" figure sanctioned is the boss of the infamous St. Petersburg „troll factory" Yevgeniy Prigozhin, who got sanctioned because of his links to the military company Wagner Group, but not because of spreading disinformation: European Union (2020).

[34] Some of the claims from the awarded TV channel include: Paedophilia is common among US politicians; gay politicians are being installed in Europe, who break children's psyche and force them to change sex; the 9/11 terror attacks in the US were in fact organised by US intelligence agencies." EUvsDisinfo (2017d); the regime also awarded pseudojournalists for "participating in the war in Syria". Not for reporting, but for participating in a war: Бойко (2016).

This US example shows another important dimension of the "Fourth line of defense"—it is crucial to investigate hostile disinformation activities. Without proper investigation, it is close to impossible to decide for a just punishment. There are several different investigations into the Russian interference in the 2016 and 2020 US Presidential elections (Ballotpedia, 2020). Some of them have already yielded criminal consequences, as shown above. In order to be able to punish these information aggressors by law, it is imperative to investigate their activities. However, despite more than a dozen elections and referenda having been attacked by the Kremlin's disinformation operations over the past few years (Kalenský, 2019a), there has not been a single similar investigation anywhere in Europe. The closest would probably be the UK parliamentary report regarding the Brexit referendum—which actually blames the government for neglecting its duties and failing to properly investigate how extensive the damage was (Piper & James, 2020).[35] This failure to investigate the activities of information aggressors is likely detrimental to limiting the willingness of other actors to employ disinformation.

7.6 Conclusion

We do not have a problem with understanding the actions of the information aggressors, the methods they are using, or with the motives they have and the degree to which these actions are menacing. Nor are there issues with understanding how to solve the disbalance between their capabilities and ours, or how to close the gap between their knowledge and ours. However, there are issues with determination; these information aggressors are determined to harm us, and we need to be able to effectively defend ourselves and our way of life, which includes a freedom of choice unaffected by professional information aggressors.

There may be a challenge to follow the advice of experts and transform their knowledge into tangible actions. Policy makers may be inclined to listen to those who underestimate information aggressors, limiting potential actions or response. It is imperative to be cognizant of the experiences of those who are on the frontline of the Kremlin's or Beijing's information aggression, such as those neighboring Russia and China. An additional challenge is sheer numbers, with information aggressors are outnumbering us—they are beating us with brute force, not just with sophistication.[36] The fact that they devote more resources also means that the gap between their knowledge and ours grows exponentially.

[35] Piper, E., & James, W. (2020). UK government failed to find out whether Russia meddled in Brexit vote: report. *reuters.com*. https://web.archive.org/web/20210511132536/https://www.reuters.com/article/us-britain-russia-idUSKCN24M15I.

[36] See e.g. Stengel, R. (2019). *Information Wars: How We Lost the Global Battle Against Disinformation and What We Can Do About It*. Atlantic Monthly Press.

It is true that the values and the way of life in democratic societies are better and stronger than those represented by the information aggressors, and we must continue to devote resources and develop strategies to contend with information aggressors.

References

Aleksejeva, N. (2020). Baltnews Latvia's audience size decreased after domain suspension. *medium.com/dfrlab.* https://web.archive.org/web/20210510191423/https://medium.com/dfrlab/baltnews-latvias-audience-size-decreased-after-domain-suspension-d636018f8a06

Anstead, N. (2021). *What do we know and what should we do about fake news?* (1st ed.). Sage Publications Ltd.

Antoniuk, D. (2020). Ukraine prolongs ban on Russian websites VKontakte, Odnoklassniki until 2023. *kyivpost.com.* https://web.archive.org/web/20210510185724/https://www.kyivpost.com/technology/ukraine-prolongs-ban-on-russian-websites-vkontakte-odnoklassniki-until-2023.html

Applebaum, A. (2019). Italians decided to fight a conspiracy theory. Here's what happened next. *washingtonpost.com.* https://web.archive.org/web/20210510093738/https://www.washingtonpost.com/opinions/global-opinions/italians-decided-to-fight-a-conspiracy-theory-heres-what-happened-next/2019/08/08/ca950828-ba10-11e9-b3b4-2bb69e8c4e39_story.html

Applebaum, A. (2021). The science of making Americans hurt their own country. *theatlantic.com.* https://web.archive.org/web/20210509095013/https://www.theatlantic.com/ideas/archive/2021/03/russia-studied-how-get-americans-make-mistakes/618328/

Aro, J. (2015a). This is what pro-Russia Internet propaganda feels like—Finns have been tricked into believing in lies. *kioski.yle.fi.* https://web.archive.org/web/20210510145150/https://kioski.yle.fi/omat/this-is-what-pro-russia-internet-propaganda-feels-like

Aro, J. (2015b). Yle Kioski investigated: This is how pro-Russia trolls manipulate Finns online—Check the list of forums favored by propagandists. *kioski.yle.fi.* https://web.archive.org/web/20210509091118/https://kioski.yle.fi/omat/troll-piece-2-english

Ballotpedia. (2020). Investigations into Russian interference in the 2016 presidential election. *ballotpedia.org.* https://web.archive.org/web/20210511132314/https://ballotpedia.org/Investigations_into_Russian_interference_in_the_2016_presidential_election

Barnes, J. E. (2018). U.S. begins first cyberoperation against Russia aimed at protecting elections. *nytimes.com.* https://web.archive.org/web/20210510183314/https://www.nytimes.com/2018/10/23/us/politics/russian-hacking-usa-cyber-command.html

Barnes, J. E. (2021). Russian interference in 2020 included influencing trump associates, report says. *nytimes.com.* https://web.archive.org/web/20210509095728/https://www.nytimes.com/2021/03/16/us/politics/election-interference-russia-2020-assessment.html

Barnes, J. E., & Goldman, A. (2020). Russia trying to stoke U.S. racial tensions before election, officials say. *nytimes.com.* http://web.archive.org/web/20210508094042/https://www.nytimes.com/2020/03/10/us/politics/russian-interference-race.html

BBCNews. (2018). Jessikka Aro: Finn jailed over pro-Russia hate campaign against journalist. *bbc.com.* https://web.archive.org/web/20210510150948/https://www.bbc.com/news/world-europe-45902496

BBCNews. (2019a). Czech pensioner jailed for terror attacks on trains. *bbc.com.* http://web.archive.org/web/20210508091914/https://www.bbc.com/news/world-europe-46862508

BBCNews. (2019b). Russia's RT banned from UK media freedom conference. *bbc.com.* https://web.archive.org/web/20210510182917/https://www.bbc.com/news/world-europe-48919085

Belsat.eu. (2019). Russian propagandist Solovyov known for demonizing west buys Como Lake villas. *naviny.belsat.eu.* https://web.archive.org/web/20210511102651/https://naviny.belsat.eu/en/news/russian-propagandist-solovyov-notorious-for-demonizing-west-buys-como-lake-villas/

Bender, B. (2019). Russia beating U.S. in race for global influence, Pentagon study says. *politico.com.* https://web.archive.org/web/20210509092808/https://www.politico.com/story/2019/06/30/pentagon-russia-influence-putin-trump-1535243

Bigg, C. (2014). Putin awards 300 journalists for 'objective' Crimea coverage. *rferl.org.* https://web.archive.org/web/20210511102832/https://www.rferl.org/a/putin-awards-journalists-objective-crimea-coverage/25373844.html

Brokes, F. (2020). Czech civil society fights back against fake news. *dw.com.* https://web.archive.org/web/20210511103817/https://www.dw.com/en/czech-civil-society-fights-back-against-fake-news/a-53758412

Bugajski, J. (2020). This is how Russia and China will exploit US elections. *thehill.com.* https://web.archive.org/web/20210510095647/https://thehill.com/opinion/national-security/495907-this-is-how-russia-and-china-will-exploit-us-elections

Burtch, A. (2018). Russia, not Russia: The tweet heard round the world. *cihhic.ca.* https://web.archive.org/web/20210510091639/https://cihhic.ca/2018/02/02/russia-not-russia-the-tweet-heard-round-the-world/

Byers, D. (2021). How Facebook and Twitter decided to take down Trump's accounts. *nbnews.com.* https://web.archive.org/web/20210510131554/https://www.nbcnews.com/tech/tech-news/how-facebook-twitter-decided-take-down-trump-s-accounts-n1254317

Cadwalladr, C. (2020). Fresh Cambridge Analytica leak 'shows global manipulation is out of control'. *theguardian.com.* http://web.archive.org/web/20200417021933/https:/www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation

Calamur, K. (2018). Some of the people Trump has blamed for Russia's 2016 election hack. *theatlantic.com.* https://web.archive.org/web/20210509100300/https://www.theatlantic.com/international/archive/2018/07/trump-russia-hack/565445/

Canada at NATO [@CanadaNATO]. (2014, August 27). *Geography can be tough: Here's a guide for Russian soldiers who keep getting lost & 'accidentally' entering #Ukraine*Twitter. Retrieved May 10, 2021, from https://web.archive.org/web/20210510091058/https://twitter.com/canadanato/status/504651534198927361?lang=en

Culliford, E. (2020). Twitter opens up data for researchers to study COVID-19 tweets. *reuters.com.* https://web.archive.org/web/20210510135809/https://www.reuters.com/article/us-health-coronavirus-twitter-data/twitter-opens-up-data-for-researchers-to-study-covid-19-tweets-idUSKBN22B2Q1

Delfi. (2020). Google presented Lithuanian initiative to world. *delfi.lt.* https://web.archive.org/web/20210509185345/https://www.delfi.lt/en/politics/google-presented-lithuanian-initiative-to-world.d?id=83550653

Dettmer, J. (2018). Britain, Russia waging fast and furious information war. *voanews.com.* https://web.archive.org/web/20210509192531/https://www.voanews.com/europe/britain-russia-waging-fast-and-furious-information-war

Dewey, C. (2014). Flight MH17's Wikipedia page edited by Russian government. *thestar.com.* https://web.archive.org/web/20210510124236/https://www.thestar.com/news/world/2014/07/21/flight_mh17s_wikipedia_page_edited_from_ip_address_associated_with_putins_office.html

DFRLab. (2021). *Weaponized: How rumors about COVID-19's origins led to a narrative arms race.* http://web.archive.org/web/20210512102042/https://www.atlanticcouncil.org/wp-content/uploads/2021/02/Weaponized-How-rumors-about-COVID-19s-origins-led-to-a-narrative-arms-race.pdf

Dizikes, P. (2018). Study: On Twitter, false news travels faster than true stories. *news.mit.edu.* https://web.archive.org/web/20210509190543/https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308

Donahoe, E. (2020). The rights and responsibilities of Internet platforms. *the-american-interest.com.* https://web.archive.org/web/20210510124835/https://www.the-american-interest.com/2020/07/10/the-rights-and-responsibilities-of-internet-platforms/

Dotto, C., & Morrish, L. (2021). Facebook says it's taking on Covid disinformation. So what's all this? *wired.co.uk*. https://web.archive.org/web/20210510183016/https://www.wired.co.uk/article/facebook-covid-disinformation

Dougherty, J. (2015). How the media became one of Putin's most powerful weapons. *theatlantic.com*. https://web.archive.org/web/20210510143237/https://www.theatlantic.com/international/archive/2015/04/how-the-media-became-putins-most-powerful-weapon/391062/

DW. (2015). Estonia launches own Russian-language TV channel. *dw.com*. https://web.archive.org/web/20210510140425/https://www.dw.com/en/estonia-launches-own-russian-language-tv-channel/a-18747088

EDRi. (2020). The EU's attempt to regulate Big Tech: What it brings and what is missing. *edri.org*. https://web.archive.org/web/20210510131147/https://edri.org/our-work/eu-attempt-to-regulate-big-tech/

EEAS. (2021). *Questions and answers about the East StratCom Task Force*. eeas.europa.eu Retrieved from https://web.archive.org/web/20210509184212/https://eeas.europa.eu/headquarters/headquarters-homepage/2116/-questions-and-answers-about-the-east-stratcom-task-force_en

Ellehuus, R. (2020). Mind the Gaps: Russian Information Manipulation in the United Kingdom. *csis.org*. https://web.archive.org/web/20210510151904/https://www.csis.org/analysis/mind-gaps-russian-information-manipulation-united-kingdom/

ERR. (2020). Sputnik ends operations in Estonia. *news.err.ee*. https://news.err.ee/1019231/sputnik-ends-operations-in-estonia

EUDisinfoLab. (2020). The EU digital strategy unpacked: How the European Union plans to tackle disinformation over the next five years. *disinfo.eu*. https://web.archive.org/web/20210510125436/https://www.disinfo.eu/advocacy/the-eu-digital-strategy-unpacked-how-the-european-union-plans-to-tackle-disinformation-over-the-next-five-years

EUDisinfoLab. (N/A). Disinformation is a small water drop that over time can hew out a stone. *disinfo.eu*. http://web.archive.org/web/20210508102932/https://www.disinfo.eu/face/disinformation-is-small-water-drop-that-over-time-can-hew-out-a-stone/

European Commission. (2019a). *Code of Practice against disinformation: Commission recognises platforms' efforts ahead of the European elections*. https://web.archive.org/web/20210510130708/https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_19_2570

European Commission. (2019b). *Latvia's decision to suspend broadcast of the Russian language channel "Rossiya RTR" complies with EU law*. https://web.archive.org/web/20210510184634/https://digital-strategy.ec.europa.eu/en/news/latvias-decision-suspend-broadcast-russian-language-channel-rossiya-rtr-complies-eu-law

European Commission. (2020a). *Coronavirus response: Fighting disinformation*. http://web.archive.org/web/20210508094900/https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation_en

European Commission. (2020b). *Shaping Europe's digital future: Eurobarometer survey shows support for sustainability and data sharing*. https://web.archive.org/web/20210509192311/https://ec.europa.eu/commission/presscorner/detail/en/IP_20_383

European Commission. (2020c). *Code of practice on disinformation*. digital-strategy.ec.europa.eu Retrieved from https://web.archive.org/web/20210510130318/https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

European Commission. (2020d). *European democracy action plan: Making EU democracies stronger*. https://web.archive.org/web/20210510152329/https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2250

European Union. (2020). COUNCIL IMPLEMENTING REGULATION (EU) 2020/1481 of 14 October 2020 implementing Article 21(2) of Regulation (EU) 2016/44 concerning restrictive measures in view of the situation in Libya. *Official Journal of the European Union*, *63*(L 341). http://web.archive.org/web/20210525125417/https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL%3A2020%3A341%3AFULL&from=EN

European Values Center. (2021). *Go vs. Blitz: An analysis of Russian-Chinese influence cooperation in Europe*. https://web.archive.org/web/20210509100656/https://europeanvalues.cz/wp-content/uploads/2021/03/go-blitz.correction.pdf

EUvsDisinfo. (2016). What's in a name? *EUvsDisinfo.eu*. https://web.archive.org/web/20210509091628/https://euvsdisinfo.eu/whats-in-a-name/

EUvsDisinfo. (2017a). Email Chains and other propaganda tools in Central and Eastern Europe. *EUvsDisinfo.eu*. https://web.archive.org/web/20210509185456/https://euvsdisinfo.eu/email-chains-and-other-propaganda-tools-in-central-and-eastern-europe/

EUvsDisinfo. (2017b). Even the strongest are in danger. *EUvsDisinfo.eu*. https://web.archive.org/web/20210509192028/https://euvsdisinfo.eu/even-the-strongest-are-in-danger/

EUvsDisinfo. (2017c). 25 Ways of combatting propaganda without doing counter-propaganda. *EUvsDisinfo.eu*. https://web.archive.org/web/20210510135936/https://euvsdisinfo.eu/25-ways-of-combatting-propaganda-without-doing-counter-propaganda/

EUvsDisinfo. (2017d). The disinformation awards. *EUvsDisinfo.eu*. https://web.archive.org/web/20210511103008/https://euvsdisinfo.eu/the-disinformation-awards/

EUvsDisinfo. (2018a). The strategy and tactics of the pro-kremlin disinformation campaign. *EUvsDisinfo.eu*. http://web.archive.org/web/20210508100826/https://euvsdisinfo.eu/the-strategy-and-tactics-of-the-pro-kremlin-disinformation-campaign/

EUvsDisinfo. (2018b). Timeline: How Russia built two major disinformation campaigns *EUvsDisinfo.eu*. https://web.archive.org/web/20210509192708/https://euvsdisinfo.eu/timeline-how-russia-built-two-major-disinformation-campaigns/

EUvsDisinfo. (2018c). In Sweden, resilience is key to combatting disinformation. *EUvsDisinfo.eu*. https://web.archive.org/web/20210509193438/https://euvsdisinfo.eu/in-sweden-resilience-is-key-to-combatting-disinformation/

EUvsDisinfo. (2019). A disillusioned democrat. *EUvsDisinfo.eu*. https://web.archive.org/web/20210510190538/https://euvsdisinfo.eu/a-disillusioned-democrat/

EUvsDisinfo. (2020a). Consequences of disinformation. *EUvsDisinfo.eu*. http://web.archive.org/web/20210508093623/https://euvsdisinfo.eu/consequences-of-disinformation/

EUvsDisinfo. (2020b). EEAS special report update: Short assessment of narratives and disinformation around the COVID-19 pandemic (Update May–November). *EUvsDisinfo.eu*. https://web.archive.org/web/20210509184731/https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic-update-may-november/

EUvsDisinfo. (2020c). Repetition and runaway selection. *EUvsDisinfo.eu*. https://web.archive.org/web/20210509193118/https://euvsdisinfo.eu/repetition-and-runaway-selection/

EUvsDisinfo. (2020d). Disinformation can kill. *EUvsDisinfo.eu*. https://web.archive.org/web/20210509193232/https://euvsdisinfo.eu/disinformation-can-kill/

EUvsDisinfo. (2021a). "My loved one thinks Bill Gates will microchip humanity". Now what? *EUvsDisinfo.eu*. https://web.archive.org/web/20210510092549/https://euvsdisinfo.eu/my-loved-one-thinks-bill-gates-will-microchip-humanity-now-what/

EUvsDisinfo. (2021b). EEAS special report update: Short assessment of narratives and disinformation around the COVID-19 pandemic (Update December 2020–April 2021). *EUvsDisinfo.eu*. https://web.archive.org/web/20210510102027/https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic-update-december-2020-april-2021/

EUvsDisinfo. (2021c). All quiet on the newsfront? Part 2. https://web.archive.org/web/20210510124401/https://euvsdisinfo.eu/all-quiet-on-the-newsfront-part-2/

EUvsDisinfo. (2021d). All quiet on the newsfront? Part 1. *EUvsDisinfo.eu*. https://web.archive.org/web/20210510130026/https://euvsdisinfo.eu/all-quiet-on-the-newsfront/

EUvsDisinfo. (2021e). Pro-kremlin propagandist—The most popular Russian "journalist" of 2020. *EUvsDisinfo.eu*. https://web.archive.org/web/20210511102516/https://euvsdisinfo.eu/pro-kremlin-propagandist-the-most-popular-russian-journalist-of-2020/

Fotyga, A. (2020). *Strengthening the East Stratcom Task Force, extending its work and turning it into a fully-fledged permanent structure within the EEAS*. europarl.europa.eu Retrieved from https://web.archive.org/web/20210509184544/https://www.europarl.europa.eu/doceo/document/P-9-2020-002103_EN.html

France24. (2020). Facebook steps up fight against virus fakery. *france24.com*. https://web.archive.org/web/20210510132119/https://www.france24.com/en/20200416-facebook-steps-up-fight-against-virus-fakery

France24. (2021). Covid-19: Putin hits back at 'strange' EU criticism of Sputnik vaccine. *france24.com*. https://web.archive.org/web/20210510100223/https://www.france24.com/en/health/20210322-putin-hits-back-after-eu-official-says-europe-has-no-need-for-sputnik-v-vaccine

Fried, D., & Polyakova, A. (2018). *Democratic defense against disinformation*. https://web.archive.org/web/20210509093133/https://www.atlanticcouncil.org/wp-content/uploads/2018/03/Democratic_Defense_Against_Disinformation_FINAL.pdf

Galeotti, M. (2015). The west is too paranoid about Russia's information war. *theguardian.com*. http://web.archive.org/web/20210508094303/https://www.theguardian.com/world/2015/jul/07/russia-propaganda-europe-america

GDI. (2021). *Russian influence disinformation: Ad-funded stories debunked by the EU External Action Service*. https://web.archive.org/web/20210511103613/https://disinformationindex.org/wp-content/uploads/2021/04/April-13-2021-DisinfoAds-Russian-influence-disinformation.pdf

Giles, K. (2016). *Handbook of Russian information warfare*. https://web.archive.org/web/20210510094029/https://www.ndc.nato.int/news/news.php?icode=995

Global Engagement Center. (2020a). *GEC Counter-Disinformation Dispatches #1*. Washington, DC: e.america.gov Retrieved from https://web.archive.org/web/20210509185046/https://e.america.gov/t/ViewEmail/i/C18DF8CD7B64226D2540EF23F30FEDED

Global Engagement Center. (2020b). *Three ways to counter disinformation*. Washington, DC. Retrieved from https://web.archive.org/web/20210510092738/https://e.america.gov/t/ViewEmail/i/E66BBA648EDAE2B02540EF23F30FEDED/2900ABE908DC4CBCE663AB054A538FBA

Global Engagement Center. (2020c). *The myth that debunking doesn't work*. Washington, DC: e.america.gov Retrieved from https://web.archive.org/web/20210509101152/https://e.america.gov/t/ViewEmail/i/8D41EB2341B3EE972540EF23F30FEDED/F2AB8F86DC5635A4AF060D6555554232

Goldberg, J. (2020). Why Obama fears for our democracy. *theatlantic.com*. https://web.archive.org/web/20210510182606/https://www.theatlantic.com/ideas/archive/2020/11/why-obama-fears-for-our-democracy/617087/

Green-Riley, N., & Stewart, C. (2020). A Clapback to Russian Trolls. *theroot.com*. https://web.archive.org/web/20210510095401/https://www.theroot.com/a-clapback-to-russian-trolls-1841932843

Grossman, S. (2020). *Russian disinformation campaigns target Africa: An interview with Dr. Shelby Grossman* [Interview]. africacenter.org. https://web.archive.org/web/20210509095305/https://africacenter.org/spotlight/russian-disinformation-campaigns-target-africa-interview-shelby-grossman/

Haag, M., & Salam, M. (2017). Gunman in 'Pizzagate' shooting is sentenced to 4 years in prison *nytimes.com*. http://web.archive.org/web/20210508083537/https://www.nytimes.com/2017/06/22/us/pizzagate-attack-sentence.html

Harwell, D., & Timberg, C. (2019). YouTube recommended a Russian media site thousands of times for analysis of Mueller's report, a watchdog group says. *washingtonpost.com*. https://web.archive.org/web/20210510124040/https://www.washingtonpost.com/technology/2019/04/26/youtube-recommended-russian-media-site-above-all-others-analysis-mueller-report-watchdog-group-says/

Helmus, T. C., Marrone, J. V., Posard, M. N., & Schlang, D. (2020). *Russian propaganda hits its mark*. https://web.archive.org/web/20210510125305/https://www.rand.org/pubs/research_reports/RRA704-3.html

Henley, J. (2020). How Finland starts its fight against fake news in primary schools. *theguardian.com*. https://web.archive.org/web/20210510144238/https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news

Henley, J. (2021). Merkel, Macron and Putin in talks on using Sputnik V jab in Europe, says Kremlin. *theguardian.com*. https://web.archive.org/web/20210510100405/https://www.theguardian.com/world/2021/mar/31/merkel-macron-and-putin-sputnik-v-vaccine-eu

Henley, J., & McIntyre, N. (2020). Survey uncovers widespread belief in 'dangerous' Covid conspiracy theories. *theguardian.com*. http://web.archive.org/web/20210508100600/https://www.theguardian.com/world/2020/oct/26/survey-uncovers-widespread-belief-dangerous-covid-conspiracy-theories

Hern, A. (2020). Twitter aims to limit people sharing articles they have not read. *theguardian.com*. https://web.archive.org/web/20210510135503/https://www.theguardian.com/technology/2020/jun/11/twitter-aims-to-limit-people-sharing-articles-they-have-not-read

Independent. (2021). Brazil regulator rejects Sputnik vaccine; Russia cries foul. *independent.co.uk*. https://web.archive.org/web/20210510100559/https://www.independent.co.uk/news/brazil-regulator-rejects-sputnik-vaccine-russia-cries-foul-brazil-russia-sputnik-russian-direct-investment-fund-rio-de-janeiro-b1838530.html

Janda, J. (2016). Czech president is Russia's Trojan Horse. *euobserver.com*. https://web.archive.org/web/20210510153251/https://euobserver.com/opinion/133789

Jeden svět na školách. (2019). *Jeden svět na školách a Youtuber Kovy - Kovyho mediální ring*. https://web.archive.org/web/20210510093348/https://www.clovekvtisni.cz/jeden-svet-na-skolach-a-youtuber-kovy-kovyho-medialni-ring-5604gp

Kalenský, J. (2019a). *Russian disinformation attacks on elections: Lessons from Europe*. E. Foreign Affairs Subcommittee on Europe, Energy, and the Environment. https://web.archive.org/web/20210509092416/https://docs.house.gov/meetings/FA/FA14/20190716/109816/HHRG-116-FA14-Wstate-KalenskJ-20190716.pdf

Kalenský, J. (2019b). A change of tactics: Blurring disinformation's source. *disinfoportal.org*. http://web.archive.org/web/20190615002208/https:/disinfoportal.org/a-change-of-tactics-blurring-disinformations-source/

Kalenský, J. (2020a). Six reasons the Kremlin spreads disinformation about the coronavirus. *medium.com/dfrlab*. http://web.archive.org/web/20210508101331/https://medium.com/dfrlab/commentary-six-reasons-the-kremlin-spreads-disinformation-about-the-coronavirus-8fee41444f60

Kalenský, J. (2020b). How the bundestag research services helped Putin's propaganda. *medium.com/dfrlab*. https://web.archive.org/web/20210509095510/https://medium.com/dfrlab/analysis-how-the-bundestag-research-services-helped-putins-propaganda-2b3972119d00

Kalenský, J. (2020c). Hitting COVID-19 disinfo websites where it hurts: Their wallets. *medium.com/dfrlab*. https://web.archive.org/web/20210511104123/https://medium.com/dfrlab/op-ed-hitting-covid-19-disinfo-websites-where-it-hurts-their-wallets-fe4a20080ad1

Kapuciánová, A. (2020). *Seznam.cz nově upozorňuje své uživatele na dezinformační weby*. https://web.archive.org/web/20210510101759/https://blog.seznam.cz/2020/12/seznam-cz-nove-upozornuje-sve-uzivatele-na-dezinformacni-weby/

Karlamangla, S. (2019). Anti-vaccine activists have doctors 'terrorized into silence' with online harassment. *latimes.com*. http://web.archive.org/web/20210508092730/https://www.latimes.com/local/california/la-me-ln-vaccine-attacks-20190317-story.html

Kasparov, G. (2015). *Winter is coming: Why Vladimir Putin and the enemies of the free world must be stopped* (Kindle ed.). PublicAffairs.

Kelion, L. (2020). Coronavirus: YouTube tightens rules after David Icke 5G interview. *bbc.com*. https://web.archive.org/web/20210510132318/https://www.bbc.com/news/technology-52198946

Keršanskas, V. (2021). *Deterring disinformation? Lessons from Lithuania's countermeasures since 2014* (Hybrid CoE Papers, Issue). https://web.archive.org/web/20210510152650/https://

www.hybridcoe.fi/wp-content/uploads/2021/04/20210427_Hybrid-CoE-Paper-6_Deterring_disinformation_WEB.pdf

Kharazian, Z., & Knight, T. (2020). Why the debunked COVID-19 conspiracy video "Plandemic" won't go away. *medium.com/dfrlab.* https://web.archive.org/web/20210510125708/https://medium.com/dfrlab/why-the-debunked-covid-19-conspiracy-video-plandemic-wont-go-away-c9dd36c2037c

Kube, C. (2020). Russia will try to meddle in 2020 U.S. election, intelligence report says. *nbcnews.com.* https://web.archive.org/web/20210510152547/https://www.nbcnews.com/politics/national-security/russia-will-try-meddle-2020-u-s-election-intelligence-report-n1134446

Laity, M. (2015). *NATO and the Power of Narrative.* In: Information at War: From China's Three Warfares to NATO's Narratives. Legatum Institute. https://web.archive.org/web/20210510140159/https://li.com/reports/information-at-war-from-chinas-three-warfares-to-natos-narratives/

Lee, Y., & Hamacher, F. (2019). Taiwan passes law to combat Chinese influence on politics. *reuters.com.* https://web.archive.org/web/20210510190358/https://www.reuters.com/article/us-taiwan-lawmaking-idUSKBN1YZ0F6

Lewandowsky, S., & Cook, J. (2020). *The conspiracy theory handbook*. https://web.archive.org/web/20210510092212/https://www.climatechangecommunication.org/wp-content/uploads/2020/03/ConspiracyTheoryHandbook.pdf

Linvill, D., & Warren, P. (2019). That uplifting tweet you just shared? A Russian troll sent it. *rollingstone.com.* http://web.archive.org/web/20200408044028if_/https:/www.rollingstone.com/politics/politics-features/russia-troll-2020-election-interference-twitter-916482/

Lipton, E., Sanger, D. E., & Shane, S. (2016). The perfect weapon: How Russian cyberpower invaded the U.S. *nytimes.com.* https://web.archive.org/web/20210510095853/https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html

LSM.lv. (2019). Latvian broadcast regulator suspends nine Russian TV channels. *eng.lsm.lv.* https://web.archive.org/web/20210510192023/https://eng.lsm.lv/article/society/society/latvian-broadcast-regulator-suspends-nine-russian-tv-channels.a339012/

LSM.lv. (2021). Latvia's broadcast regulator bans Russian channel for a year. *eng.lsm.lv.* https://web.archive.org/web/20210510184814/https://eng.lsm.lv/article/features/media-literacy/latvias-broadcast-regulator-bans-russian-channel-for-a-year.a392026/

Lucas, E., & Pomeranzev, P. (2016). *Winning the Information War*. https://web.archive.org/web/web/20210509093947/https://li.com/wp-content/uploads/2016/08/winning-the-information-war-full-report-pdf.pdf

Mackintosh, E. (2019). Finland is winning the war on fake news. What it's learned may be crucial to Western democracy. *edition.cnn.com.* https://web.archive.org/web/20210510143759/https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/

Martin, D. A., Shapiro, J. N., & Ilhardt, J. G. (2020). *Trends in online influence efforts.* E. S. o. C. Project. https://web.archive.org/web/20210509094317/https://drive.google.com/file/d/18QIENHZslNIoKvOu72iEjG6RgWL1Dww_/view

Mayer, J. (2018). How Russia helped swing the election for Trump. *newyorker.com.* http://web.archive.org/web/20210508093449/https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump

McCauley, K. (2016). *Russian influence campaigns against the west: From the Cold War to Putin* (Kindle ed.). CreateSpace Independent Publishing Platform.

McGuiness, D. (2016). Russia steps into Berlin 'rape' storm claiming German cover-up. *bbc.com.* http://web.archive.org/web/20210508093815/https://www.bbc.com/news/blogs-eu-35413134

McTague, T. (2019). Britain's secret war with Russia. *theatlantic.com.* http://web.archive.org/web/20210508101203/https://www.theatlantic.com/international/archive/2019/12/britain-russia-nato-disinformation/602836/

Meduza. (2016). Russian court convicts man of 'falsifying history' for saying the USSR shares responsibility for starting WWII. *meduza.io.* https://web.archive.org/web/20210510094524/https://meduza.io/en/news/2016/07/01/russian-court-convicts-man-of-falsifying-history-for-saying-the-ussr-shares-responsibility-for-starting-wwii

Meister, S. (2016). *The "Lisa case": Germany as a target of Russian disinformation* (NATO Review, Issue. nato.int. https://web.archive.org/web/20210509193909/https://www.nato.int/docu/review/articles/2016/07/25/the-lisa-case-germany-as-a-target-of-russian-disinformation/index.html

Michail, N. (2019). 'Overwhelmingly positive': Chile's food regulations are changing the country's eating habits. *foodnavigator-latam.com.* https://web.archive.org/web/20210510125104/https://www.foodnavigator-latam.com/Article/2019/02/20/Chile-s-food-regulations-are-changing-the-country-s-eating-habits

Mierzyńska, A. (2021). Rosyjska propaganda dezinformuje, manipulując wpisami na polskich portalach informacyjnych. *oko.press*. https://web.archive.org/web/20210510150637/https://oko.press/rosyjska-propaganda-dezinformuje-manipulujac-wpisami-na-polskich-portalach/

Ministerstvo vnitra ČR. (2016). *Centre against terrorism and hybrid threats*. https://web.archive.org/web/20210509193627/https://www.mvcr.cz/cthh/clanek/centre-against-terrorism-and-hybrid-threats.aspx

MSB. (2019). *Countering information influence activities—A handbook for communicators*. http://web.archive.org/web/20200407131305/https:/www.msb.se/RibData/Filer/pdf/28698.pdf

NATO. (2020). *NATO—Russia. Setting the record straight*. nato.int. Retrieved from https://web.archive.org/web/20210509183926/https://www.nato.int/cps/en/natohq/115204.htm

Newman, L. H. (2020). Russia is learning how to bypass Facebook's disinfo defenses. *wired.com.* https://www.wired.com/story/russia-ira-bypass-facebookdisinfo-defenses/

NFNZ. (2020). *Rating médií*. https://web.archive.org/web/20210510101633/https://www.nfnz.cz/rating-medii/

Nimmo, B. (2018). Question that: RT's military mission. *medium.com/dfrlab*. https://web.archive.org/web/20210510143518/https://medium.com/dfrlab/question-that-rts-military-mission-4c4bd9f72c88

Nimmo, B. (2020). *The breakout scale: Measuring the impact of influence operations*. https://web.archive.org/web/20210509190406/https://www.brookings.edu/wp-content/uploads/2020/09/Nimmo_influence_operations_PDF.pdf

O'Kane, C. (2019). Russian trolls fueled anti-vaccination debate in U.S. by spreading misinformation on Twitter, study finds. *cbsnews.com*. http://web.archive.org/web/20210508093145/https://www.cbsnews.com/news/anti-vax-movement-russian-trolls-fueled-anti-vaccination-debate-in-us-by-spreading-misinformation-twitter-study/

Owen, J. (2018). Britain behind the curve in fighting fake news, admits government comms chief. *prweek.com*. https://web.archive.org/web/20210509183652/https://www.prweek.com/article/1484791/britain-behind-curve-fighting-fake-news-admits-government-comms-chief

Partin, W. (2020). What if modern conspiracy theorists are altogether too media literate? *theoutline.com*. https://web.archive.org/web/20210510144733/https://theoutline.com/post/8509/revolution-q-conspiracy-theorists-media-literacy

Pettersen, T. (2016). Sputnik closes Nordic language services. *thebarentsobserver.com*. https://web.archive.org/web/20210509090934/https://thebarentsobserver.com/en/society/2016/03/sputnik-closes-nordic-language-services

Piper, E., & James, W. (2020). UK government failed to find out whether Russia meddled in Brexit vote: report. *reuters.com*. https://web.archive.org/web/20210511132536/https://www.reuters.com/article/us-britain-russia-idUSKCN24M15I

Pogrund, G. (2020). Army spies to take on antivax militants. *thetimes.co.uk*. https://web.archive.org/web/20210509185949/https://www.thetimes.co.uk/article/army-spies-to-take-on-antivax-militants-mfzsj66w2?region=global&--xx-meta=denied_for_visit%3D0%26visit_number%3D0%26visit_remaining%3D0%26visit_used%3D0&--xx-mvt-opted-out=false&--xx-uuid=86f34e229f7fb279b9fc1f40c7280179&ni-statuscode=acsaz-307

Pomerantsev, P., & Weiss, M. (2014). *The menace of unreality: How the kremlin weaponizes information, culture and money*. https://web.archive.org/web/20210510101234/https://imrussia.org/media/pdf/Research/Michael_Weiss_and_Peter_Pomerantsev__The_Menace_of_Unreality.pdf

President Joseph R. Biden, J. (2021). *Interim national security strategic guidance*. Washington, DC: whitehouse.gov Retrieved from https://web.archive.org/web/20210510094715/https://www.whitehouse.gov/wp-content/uploads/2021/03/NSC-1v2.pdf

RFE/RL. (2016). Medvedev's awkward Crimea moment: 'There's just no money. But you take care!'. *rferl.org*. https://web.archive.org/web/20210510093224/https://www.rferl.org/a/russia-medvedev-crimea-visit-no-money-social-media-pensioner/27754644.html

RFE/RL. (2017). OSCE calls on Estonia to reconsider ban on three Russian reporters. *rferl.org*. https://web.archive.org/web/20210510182753/https://www.rferl.org/a/osce-desir-calls-on-estonia-reconsider-ban-russian-reporters-russia-today-sputnik/28706515.html

RFE/RL. (2019). Lithuania expels chief editor of Sputnik's local branch. *rferl.org*. https://web.archive.org/web/20210510183835/https://www.rferl.org/a/lithuania-expels-chief-editor-of-sputnik-local-branch/29968909.html

RFE/RL. (2020a). COVID-19 Disinformation from Russia, China more influential than domestic news sources, study finds. *rferl.org*. https://web.archive.org/web/20210509190708/https://www.rferl.org/a/covid-19-disinformation-from-russian-chinese-media-more-influential-than-domestic-news/30696566.html

RFE/RL. (2020b). Bulgaria charges pro-Russia politician with spreading false information about coronavirus. *rferl.org*. https://web.archive.org/web/20210510190047/https://www.rferl.org/a/bulgaria-charges-pro-russia-politician-with-spreading-false-information-about-coronavirus/30518495.html

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare* (Kindle ed.). Profile Books.

Sanger, D. E., & Perlroth, N. (2019). U.S. escalates online attacks on Russia's power grid. *nytimes.com*. https://web.archive.org/web/20210510183442/https://www.nytimes.com/2019/06/15/us/politics/trump-cyber-russia-grid.html

Schearf, D. (2017). Baltics' Russian media use online humor to combat propaganda. *voanews.com*. https://web.archive.org/web/20210510093108/https://www.voanews.com/europe/baltics-russian-media-use-online-humor-combat-propaganda

Schoen, F., & Lamb, C. J. (2012). *Deception, disinformation, and strategic communications: How one interagency group made a major difference*. Strategic Perspectives, Issue. N. D. U. Press. http://web.archive.org/web/20210508102102/https://ndupress.ndu.edu/Portals/68/Documents/stratperspective/inss/Strategic-Perspectives-11.pdf

Schultz, T. (2017). Why the 'fake rape' story against German NATO forces fell flat in Lithuania. *dw.com*. https://web.archive.org/web/20210509194301/https://www.dw.com/en/why-the-fake-rape-story-against-german-nato-forces-fell-flat-in-lithuania/a-37694870

Shandra, A., & Seely, R. (2019). *The Surkov Leaks. The inner workings of Russia's hybrid war in Ukraine*. https://web.archive.org/web/20210510151351/https://rusi.org/sites/default/files/201907_op_surkov_leaks_web_final.pdf

Shutov, R. (2020). The Kremlin's reputational losses do not mean that the West is winning the information war. *medium.com/dfrlab*. https://web.archive.org/web/20210509191224/https://medium.com/dfrlab/op-ed-the-kremlins-reputational-losses-do-not-mean-that-the-west-is-winning-the-information-war-b207c248b1d2

Snegovaya, M., & Watanabe, K. (2021). *The Kremlin's social media inside the United States: A moving target*. F. R. Foundation. https://web.archive.org/web/20210509090339/https://www.4freerussia.org/the-kremlin-s-social-media-influence-inside-the-united-states-a-moving-target/

Soldatov, A. (2021). *Ruský expert na tajné služby: Vrbětice jsou neskutečný příběh. Ukazují, v co se ruská rozvědka proměnila* [Interview]. irozhlas.cz. https://web.archive.org/web/20210510182329/https://www.irozhlas.cz/zpravy-svet/podcast-vinohradska-12-vrbetice-andrej-soldatov-rusko-tajne-sluzby_2104300600_mpa

Sostav. (2020). ТОП-30 крупнейших рекламодателей России 2020. *sostav.ru*. http://web.archive.org/web/20210505161435/https:/www.sostav.ru/publication/top-30-advertizers-2020-43324.html

Spectator. (2019). Editor-in-chief of conspiracy magazine pleaded guilty. *spectator.sme.sk*. https://web.archive.org/web/20210510190138/https://spectator.sme.sk/c/22285273/editor-in-chief-of-a-conspiracy-magazine-pleaded-guilty.html

Spurný, M. (2019). *Výzkum veřejného mínění k problematice dezinformací*. M. v. ČR. https://web.archive.org/web/20210509190947/https://www.mvcr.cz/cthh/clanek/vyzkum-verejneho-mineni-k-problematice-dezinformaci.aspx

STEM. (2021). *COVID-19 and conspiracy spreaders. Public attitude's survey in the Czech Republic*. http://web.archive.org/web/20210508095052/https://www.stem.cz/wp-content/uploads/2021/03/STEM_Disinformation-and-Conspiracies_GENERAL-POPULATION-Mar-26-2021.pdf

Stengel, R. (2019). *Information wars: How we lost the global battle against disinformation and what we can do about it*. Atlantic Monthly Press.

Swab, A. J. (2020). Baltic States offer insights into how to counter post-election disinfo. *medium.com/dfrlab*. https://web.archive.org/web/20210510144348/https://medium.com/dfrlab/op-ed-baltic-states-offer-insights-into-how-to-counter-ongoing-election-disinfo-23e67b76ead

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition, 9*(3), 286–299. https://web.archive.org/web/20210509101408/https://www.sciencedirect.com/science/article/pii/S2211368120300516

Tanner, J. (2020). Latvia bans Russian television channel RT. *apnews.com*. https://web.archive.org/web/20210510191837/https://apnews.com/article/international-news-entertainment-television-technology-russia-0c6cf24233c7097d4d7bd89697894a60

Taylor, J. (2020). Twitter deletes 170,000 accounts linked to China influence campaign. *theguardian.com*. https://web.archive.org/web/20210510135654/https://www.theguardian.com/technology/2020/jun/12/twitter-deletes-170000-accounts-linked-to-china-influence-campaign

Thomsen, J. (2017). Ukraine uses 'Simpsons' gif in Twitter fight with Russia. *thehill.com*. https://web.archive.org/web/20210510091812/https://thehill.com/blogs/blog-briefing-room/335597-ukraine-trolls-russia-with-the-simpsons-gif

Tokariuk, O. (2021). *Battle of narratives: Kremlin disinformation in the Vitaliy Markiv case in Italy*. https://web.archive.org/web/20210510101100/https://drive.google.com/file/d/1sGK2lLqN46MVMN6qFKvRvJ4buJP1Fu3b/view

U.S. Department of Treasury. (2021). *Treasury escalates sanctions against the Russian Government's attempts to influence U.S. elections*. https://web.archive.org/web/20210511104301/https://home.treasury.gov/news/press-releases/jy0126

UAWire. (2020). Latvia denies entry to Russian journalist. *uawire.org*. https://web.archive.org/web/20210510184037/http://www.uawire.org/latvia-denies-entry-to-russian-journalist

UK Government. (2019). *RESIST counter-disinformation toolkit*. http://web.archive.org/web/20200407130817/https:/gcs.civilservice.gov.uk/news/resist-counter-disinformation-toolkit-launched-today/

UK Government. (2021). *Global Britain in a competitive age: The integrated review of security, defence, development and foreign policy*. gov.uk: gov.uk Retrieved from http://web.archive.org/web/20210508105143/https://www.gov.uk/government/publications/global-britain-in-a-competitive-age-the-integrated-review-of-security-defence-development-and-foreign-policy/global-britain-in-a-competitive-age-the-integrated-review-of-security-defence-development-and-foreign-policy

UNIAN. (2020). Russian propagandists denied entry, banned from Ukraine. *unian.info*. https://www.unian.info/society/10884716-russian-propagandists-denied-entry-banned-from-ukraine.html

Vandiver, J. (2014). SACEUR: Allies must prepare for Russia 'hybrid war'. *stripes.com*. https://web.archive.org/web/20210509100052/https://www.stripes.com/news/saceur-allies-must-prepare-for-russia-hybrid-war-1.301464

Velychko, L. (2020). How pro-Kremlin Telegram channels influence Ukrainian parliamentary decisions. *medium.com/dfrlab*. https://web.archive.org/web/20210509185623/https://medium.

com/dfrlab/how-pro-kremlin-telegram-channels-influence-ukrainian-parliamentary-decisions-791ac939cdd

Vilmer, J.-B. J., & Charon, P. (2020). Russia as a hurricane, China as climate change: Different ways of information warfare. *warontherocks.com.* https://web.archive.org/web/20210509100400/https://warontherocks.com/2020/01/russia-as-a-hurricane-china-as-climate-change-different-ways-of-information-warfare/

Visions, S. (2020). *Russia leading from behind: The origins of Covid-19 disinformation.* https://web.archive.org/web/20210509095845/https://semantic-visions.com/resource/russia-leading-from-behind-coronavirus-focused-case-study-of-cross-border-disinformation-spread

Wang, Y. C. (2019). EU court backs Lithuania in TV hate speech case. *politico.eu.* https://web.archive.org/web/20210510183707/https://www.politico.eu/article/eu-court-backs-lithuania-in-tv-hate-speech-case/

Weiss, M., & Goldsmith, J. (2021). Syria chemical-attack deniers admit links to WikiLeaks and Russia. *thedailybeast.com.* https://web.archive.org/web/20210509092956/https://www.thedailybeast.com/syria-chemical-attack-deniers-admit-links-to-wikileaks-and-russia

Wesslau, F. (2016). *Putin's friends in Europe.* https://web.archive.org/web/20210509094755/https://ecfr.eu/article/commentary_putins_friends_in_europe7153/

WHO. (2020, September 23). *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation.* http://web.archive.org/web/20210316122039/https:/www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation

Yang, O. (2020). Taiwan's battle against COVID-19 disinformation. *medium.com/dfrlab.* https://web.archive.org/web/20210510092016/https://medium.com/dfrlab/taiwans-battle-against-covid-19-disinformation-eaf76bf57d14

Бойко, А. (2016). В Москве наградили лучших военных журналистов. *kp.ru.* https://web.archive.org/web/20160418144138/https:/www.kp.ru/daily/26518.7/3534589/

Українська правда. (2021). ЗЕЛЕНСЬКИЙ ЗАБЛОКУВАВ КАНАЛИ МЕДВЕДЧУКА: РНБО НАКЛАЛА САНКЦІЇ. *pravda.com.ua.* https://web.archive.org/web/20210510185853/https://www.pravda.com.ua/news/2021/02/2/7282097/

Chapter 8
Are You Seeing What I Am Seeing? Ensuring Data Relevance for Online Information Environment Assessments

Arild Bergh

Abstract Social media analysis has gone from an arcane, minority art to a mainstream necessity in a few short years. Following on from interference in US elections and attacks on democracy campaigners, the COVID-19 crisis has shown that social media is still an easy option for actors who want to disrupt democracies by pushing disinformation and amplifying misinformation. Although there are numerous tools and services available that offer to analyse social media, defence and general government strategic communication staff can find it difficult to evaluate their quality. Of particular concern is the oblique nature of the underlying data which can skew assessments quite dramatically. This chapter will use the experiences of Norwegian strategic communication personnel as a lens to highlight problems one might encounter when working with social media analysis. This approach grounds the recommendations and suggestions for practical solutions on how to evaluate the relevance of dataset used in online information environment assessments.

8.1 Introduction

The emergence of social media as a vehicle for large-scale disinformation campaigns, targeted influence operations and general misinformation over the past decade has made social media a problem for decision makers in defence and total defence (Bergh, 2019). Early on in the COVID-19 pandemic it was noted how the pandemic was accompanied by an "infodemic" (Cinelli et al., 2020; Richtel, 2020)—the barrage of bad information that made the handling of the pandemic more difficult and, in extreme consequences, caused the loss of life (Islam et al., 2020). This infodemic takes place within similar social media based efforts by state actors to affect elections, skew public debate in their favour and increase polarisation in other countries (see for example Hamilton & Ohlberg, 2020). There is thus the need for a continuously updated overview over what happens in social media through what this book refers

A. Bergh (✉)
Norwegian Defence Research Establishment (FFI), Kjeller, Norway
e-mail: Arild.Bergh@ffi.no

R. Gill and R. Goolsby (eds.), *COVID-19 Disinformation: A Multi-National, Whole of Society Perspective*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-94825-2_8

to as “online information environment assessment”. The bedrock of such assessment is the data analysed, hence this chapters focus on the data and not the analysis.

This chapter summarises the lessons learned from the author’s support to the Norwegian “DepStrat” group. This is an interdepartmental strategic communications group led by the Ministry of Defence that has emerged as a key actor in terms of assessing and handling dis- and misinformation spread through social media. The group had been in the planning stages for some time prior to the COVID-19 pandemic. However, it received a baptism of fire in the early weeks of the pandemic as the need for an online information environment assessment became pressing. As analyses of the infodemic were in high demand without an established pipeline for the purpose, the group explored a number of approaches in rapid succession.

The key takeaway discussed in this chapter is a frequent fallacy in the initial approach of governmental or defence actors—it is all too easy perceive social media as just another data source, similar to other sources of data. This assumption leads to the belief that social media data are uniform and that good assessment is mainly a question of quantity (and more is better). The purpose of this chapter is to help communication staff and decision makers to avoid this fallacy. This chapter will outline what makes social media data different from other data sources; how this affects collection and analysis of data; how to determine one’s actual data requirements for an online information environment assessment task and how to evaluate the relevance of available data to that task. The focus here is on understanding the basis on which analysis is built. Data relevance is important. Analysts must have a strong grasp of this problem or they risk trying to navigate the information landscape with the wrong maps.

8.2 Definitions

Throughout this chapter, *disinformation* will be used as a generic term that covers fake news, deliberate influence operations and other forms of false information. Misinformation, incorrect information that is shared by people who believe it to be correct is often difficult to differentiate from disinformation and the effects can be similar when applied to the problem of responding to an “infodemic.” In this chapter therefore, disinformation is used as a blanket term. *Data base/data basis* (as opposed to database) refers to the raw data that is analysed. *Analysis/analytics* generally refers to any assessment of the online information environment.

8.3 What Makes Data from Social Media Different

Most of us have at some point been asked to respond to a questionnaire—perhaps an online survey of favourite news sites or telephone polls asking about voting intentions. The type of data being gathered is predetermined and fairly limited in what it says about a person. The number of responses typically run into a few hundred or lower thousands. More in-depth and contextualised data are sought in longer, semi-structured person-to-person interviews, for instance, research on gender issues at work. The actual data is more detailed, but still about a fixed area of someone's life and the typical number of responses reach a few dozen at the most.

The data from these approaches are narrow snapshots of a small part of a person's life, devoid of information about how it relates to other parts of life and with little overall context. The data is gathered in rather small quantities, is relatively time and money consuming to collect and requires considerable planning to get right.

These traditional data collection methods are quite limited compared to data from social media. Social media provides an abundance of data that is comparatively cheap to gather in large quantities (Johnes, 2012). If data is gathered over time from the same sources one can observe changes to someone's (online) life almost in real time. The data will provide all kinds of linkages to other people, but also to thoughts, ideas, and interests. It can also be collected relatively ad-hoc without specifying what one is looking for, and then analysed retrospectively. For instance, one could collect all tweets for a given day and later see how many used a particular hashtag in different countries.

It is the latter aspect that can be the most seductive, but also the most problematic, when collecting data from social media. Social media data does not require a carefully planned data collection to find out what *sample A* thinks about *specific issue B*, the way that traditional data collection methods require. The data available are so vast and varied that one will easily find "something about anything". It is as if the data wraps itself around any issue one wants to investigate to fit in with what is asked of it. Data collection methods and criteria must be explicit: in analysis, one must be able to distinguish the methods, keywords, and other filtering methods in the sample to establish provenance. Otherwise, when the basis for the data is of unknown provenance (because one does not know or understand the original selection criteria) one might end up with decision-based evidence making rather than evidence-based decision making. Failure to understand provenance can lead to a number of errors, such as identifying a particular hashtag as very important, but not accounting for the fact that this hashtag was one of the search terms in collecting the sample. This can inflate the number of tweets using that hashtag in the sample. This is not a problem that is unique to social media information environment assessment. However, social media analysis can be very convincing due to the scale of the data base. Therefore, ensuring that the data is *relevant* for the problem under consideration is a key consideration.

The remainder of this chapter will explain the lessons learned from how Norway's DepStrat group encountered these pitfalls during their response to COVID-19 and discuss how others can avoid them by evaluating the data base used by external providers of social media analysis. This can improve the validity of the assessment of online disinformation in times of crises or conflicts.

8.4 Obstacles to Data Relevancy

As the DepStrat group started working to obtain data and analyses for an assessment of the disinformation surrounding COVID-19 there were no lack of commercial actors that could supply social media data, analytics tools and complete analyses. However, the (rapidly fluctuating) requirements of DepStrat were not mainstream, and the basis for a data collection was often oblique. The needs of commercial industry to track social media are not the same as the needs for tracking virulent disinformation. Existing commercial sources vary in their ability to provide appropriate services for these tasks. When DepStrat started examining the different services, they found that commercial services often promoted their capabilities to quickly finding needles in a haystack that looked impressively large. There was little discussion as to what that haystack consisted of, as if all social media were inherently the same and quantity was the key criteria of relevance. DepStrat had to take a closer look at the marketplace to understand how it functions and what it offers. This examination exposed three key issues that affected the data base: (1) services handling of data protection laws and social media *terms of service* (TOS) and how this limited access to important data; (2) the complex supply chains used to obtain data; and (3) the reliance on oversimplification to provide automated graphical representation to provide summary analysis of results.

8.4.1 Social Media Data Collection Frictions

The properties of social media data that make it valuable for research purposes also make it sensitive from a data protection perspective. Different countries can have very different rules for what data can be collected and how it can be analysed. However, for EU and EFTA countries the General Data Protection Rules (GDPR) from 2018 are the core regulations to relate to (EU, 2018). This is an important democratic safeguard that sets clearly defined limits to the data that can be collected. It is beyond the scope of this chapter to evaluate data protection issues in any detail, suffice to say that in countries with good protection of data and citizens privacy there are a number of hurdles to clear before one can collect data from social media (Greene et al., 2019; Maldoff, 2016).

Social media platforms usually feature complex TOS that can also cause confusion when planning or evaluating data collection. These terms are often focused on giving the social media platforms the possibility of taking legal action against scammers or stopping competing companies from using their data. However, collecting data from openly available pages such as public Facebook groups, so-called "scraping", is often legal (depending on jurisdiction), as long as one follows relevant local regulations for protecting personal information. In Norway there are a number of actors who have collected data through web scraping for research or business (Bailey, 2019; Berg & Frivold, 2017; Hageskal & Ueland, 2018; Prisjakt, 2019). There are also examples of researchers who have chosen to break the TOS by creating accounts for fictional users for the purpose of obtaining relevant data for important assessments in the fight against disinformation. For example, NATO's Strategic Communications Center of Excellence in Riga examined how soldiers could be manipulated using a fake Facebook public group. The findings highlighted the challenges that could arise for military decision-makers in Latvia in connection with soldiers' use of social media. This effort contributed to major changes in the military's rules for smartphone use (Bay & Biteniece, 2019). This had no legal repercussions for the researchers.

The level of technical efforts required to harvest data from different sources can also affect the data collection. Many services and tools focus on Twitter because Twitter has a simple to use interface for querying and downloading data (Morstatter et al., 2013). When presenting findings from such datasets it is often discussed as "social media" when in fact it is a single, specific platform. This can have a huge effect on the relevance of the data; in Norway only 8% use Twitter daily whereas 67% use Facebook (Ipsos, 2020). Twitter analyses will therefore be less appropriate to Norwegian information environment assessments than an analysis of relevant Facebook groups. Similar differences will exist in other countries.

8.4.2 The Supply Chain's Effect on Data Relevance

Given the frictions outlined above there are considerable incentives for analysts, whether commercial actors or in-house, to rely on other, specialised suppliers rather than collect their own data. These suppliers include the social media platforms themselves (e.g. Twitter), third party data brokers or services that give access to data and analysis tools (the Facebook owned CrowdTangle is one such tool) or researchers sharing datasets they have collected. Thus there may be several gatekeepers and intermediaries whose decisions and data treatment affect the data's relevance to the issues one wants to explore. When DepStrat started looking into the source of data being used for analysis it was clear that different providers often used the same subcontractors to get hold of data. This market for data creates an opaque supply chain where data from many different sources is used for a range of analytical purposes.

Consequently, data may have passed through several different and fluctuating selection criteria as they are collected into datasets. The supply chain is rarely very explicit in discussing these criteria, instead is focuses on the end product, their analysis of the data. There are numerous questions one should discuss here: For example, what are the data collectors doing to distinguish bots from real users (if anything)? Which social media platforms and groups does the dataset include? Are there any default limitations? CrowdTangle for example, only collects data from Facebook groups with more than 100,000 likes. This is a criterion likely to exclude most Norwegian groups due to the population size. Which data has been filtered out by the original social media platform, and why? Twitter, for instance, has different access methods that provide different selections of data (Kim et al., 2020). Equally important is whether the data collection criteria have changed over time? For instance, have some hashtags not included in the beginning been added later? This is a common problem as hashtags evolve, particularly in crisis situations.

At the end of the supply chain, different commercial providers will have different business models that affect their data acquisition and types of analysis. If their core focus is advertising, the service may look at brands' reputation and ignore aspects that are important to information environment assessments, such as narratives and the use of bots to boost post interactions. Others may focus on elections and ignore longer term international interference on global issues like the COVID-19 pandemic.

Finally, not all datasets can be used by all customers, for various reasons. CrowdTangle specifically prohibits "government affiliated institutions, such as national research institutions" from using their Facebook data services. The result can be a skewed dataset with holes in it, holes that are not immediately evident in a social media information environment assessment (CrowdTangle, 2021). With 67% of the Norwegian population using Facebook daily, this posed problems in assessing how concerned DepStrat should be about COVID-19 disinformation on that platform. Although Facebook, like other platforms, eventually took steps to limit the spread of this type of disinformation, it was difficult to assess the efficacy of their methods.

8.4.3 Graphics, Not Data

The emergence of big data, of which social media data is one type, has created a range of innovative visualisation techniques that aim to make the findings of these massive datasets more intuitively understandable. Unfortunately, sometimes this results in a focus on attention-grabbing visualisation, such as social network maps, and less on the underlying data quality. These vanity graphics may focus more on basic summaries rather than deeper insights, often lacking the ability to filter the data or perform deeper analysis. Such graphics can be misleading as big data samples invariably contain massive amounts of "noise"—irrelevant data that happen to get caught up in the sample. This can obscure or even eclipse the relevant data in the sample. However, to an untrained eye they can look very convincing. This aspect again points to the importance of being able to assess the underlying datasets in any given situation.

For example, if one includes "corona" in the search criteria in Twitter or reddit, one would receive a strong influx of posts about Corona beer. Some disinformation did surround Corona beer in the beginning of the pandemic, but later this term would produce largely jokes and memes that were irrelevant to the task of identifying important disinformation themes and narratives. This data would cause the graphics from some commercial services to fail to deliver a good visualization of relevant data unless one could drill down, assess that data and filter it from the sample.

8.5 Determining Data's Relevance to Information Environment Assessment

The unique nature of social media data and the obstacles to a carefully controlled data collection makes it important to carefully question the underlying data when considering different providers of datasets or data analyses. This section will provide some key pointers on how to develop relevant criteria, and how to use these in a checklist to evaluate the relevance of data for information environment assessment drawn from DepStrat's experience during the COVID-19 pandemic.

Manual Analysis

Anyone can log on to a social media platform and start searching for content. Then why bother to collect data and analyse this separately? Although a manual approach can be useful to see if particular groups or influential profiles discuss a particular topic or narrative, it can easily lead one astray.

Social media algorithms (automated software that solve a given task) are set up to always try to find content that a user is interested in (Bergh, 2019, pp. 22–23). Thus, if one spends considerable time searching for "vaccine side effects" the algorithms observe this and thereafter select posts that is assumed to appeal to ant-vaccine users. In other words, one will see more and more content equal to what one has already looked at, giving the impression that there is a lot of that type of content. Not because it **is** more, but because the algorithms select it in preference to other content.

Thus, there is a definite need to analyse large datasets, which necessitates a data collection.

8.6 Determine Data Requirements for a Given Situation

In the early stages of the pandemic, when the DepStrat group attempted to gather data to assess what happened in social media around COVID-19, the group's first impulse was to look at commercial suppliers of analyses in this field. There they encountered a novel problem: how to evaluate the many products offered. Commercial services heavily promoted the quantity of data being analysed, but many did not disclose what the criteria for selecting the data had been. When someone is under pressure to produce knowledge for decisionmakers, it can be tempting to select something familiar or something that looks impressive.

The first lesson learned: develop the key data selection criteria internally, before consulting outside resources for the data. This enables a better evaluation of the underlying data being used in the analytics, tools, and visualizations offered.

Disinformation is always aimed at someone—be it a group, a personality type or an entire country. The disinformation actors also need to distribute or promote content. This leads to two core questions when collecting data from social media: (a) Who are likely to be targeted by, or seek out, disinformation; and (b) what are the probable topics and narratives that will be used to spread disinformation?

In the pandemic, health-related discourses would be the most likely vector for promoting disinformation. In some countries, political discourses became a prominent vector. Celebrities and other influencers can play important roles in disseminating disinformation to their follower communities. Propagandists may even target celebrities and influencers as they have done before. Finding a celebrity to promote a talking point can ease the way to target large communities. During the early phases of COVID-19, alternative health topics, including such mainstream topics as yoga, were targeted on Instagram and Twitter to inject disinformation in groups that were already likely to consider alternative perspectives on a health-related concern.

With the tentative answers to questions about targets and topics one can start to define the data required. This in turn allows datasets to be evaluated in terms of their relevance for the information environment assessment that will be undertaken.

With the target groups in mind, one needs to find out which social media platforms they use. This could be particular Facebook groups, niche social networks like Parler or regional/national social media sites, such as VK (previously vKontakte, Вконтакте) in Russia. There will also be national differences, for example in Norway some 24% of the population use YouTube daily vs 51% in the USA (Auxier & Anderson, 2021; Ipsos, 2020).

One also need to consider languages that can be used to spread disinformation to the target groups. This pertains to both collection and analysis of data. Can for instance machine learning models be used reliably for sentiment analysis? Many models have been trained on English text and will be less accurate for other languages. Most sentiment analysis tools fail to provide the user with access to the underlying data that were used to train their software to determine what, for example, was negative or positive. The user can't look at the data so they can't know if the score is based on highly relevant posts, or a mishmash of irrelevant and relevant posts. An example might be "Corona beer is the bomb!" which could be classed as a negative sentiment, as well being irrelevant, since "bomb" often gets classed as a negative word.

The topics that may be deployed for disinformation is then used to develop a list of key words and phrases that are applied to select social media posts for collection. In the case of the DepStrat group's work on COVID-19 disinformation these phrases were chosen initially:

> *[variants of COVID-19 names]; measures; quarantine; [names of various government department and agencies]; face mask; vaccine; [names of various spokespeople and prominent public health personnel]; social distancing.*

Such a list will obviously not be static, as a crisis develops, so will the list. Over time different terms enter the public consciousness or certain aspects of a crisis becomes more amenable to disinformation. It is therefore important that the initial list is quite broad, so it is possible to go back later and use previously collected data for new analyses to detect changes and trends over time.

This leads us to the final criteria; the period that the data needs to cover. Should it start from today? Or from when disinformation efforts are likely to have started? Or from a time before the crisis? If the data collection starts from scratch today there is no context. What level of bad information is the norm, perhaps because of genuine misunderstandings or as part of ordinary, casual banter. Datasets that go further back in time can be used to develop a baseline of normal social media activity that can tell us something about current levels of disinformation.

In sum, by defining potential targets, pertinent topics and proper time periods one is in a good position to evaluate how well a dataset fits the information environment assessment that will be performed. This is also a solid step towards evidence-based decision making.

8.7 A Social Media Data Checklist

Based on the experience of FFI's work with the DepStrat group, some key points have emerged with regard to evaluating commercial services as data suppliers for information environment assessment during crises like COVID-19. This is not an exhaustive list but it provides a solid starting point for non-experts—those who need an analysis of social media to detect disinformation but are not social media/big data analysts themselves.

1. Do the data cover relevant sources such as:
    a. The main social media platform(s) used in the country/countries being researched;
    b. The main social media platform(s) used by potential target groups;
    c. Social media groups on social media platforms used by potential target groups;
    d. Specific social media profiles that are known to spread disinformation?
2. Are there any meta-criteria that limits the data that has been collected, e.g.:
    a. Filtering of spam (such data can be used to detect bot networks);
    b. Size of social media groups on social media platforms (focus on too large groups can hide highly targeted disinformation activities);
    c. Exclusions based on (local) data protection laws;
    d. Do any of the data suppliers/analysts have an overarching business model or focus area that would limit the data being collected? Do they have experience or expertise in disinformation?
3. Do the original criteria for the data collection include, **or at least not exclude**:
    a. Topics that are deemed relevant and operationalised through keywords and hashtags;

   b. If relevant keywords have been specified, are there other criteria that could have limited the selection?

4. What is the fit between the period the data collection covers and the time when disinformation most likely was spread?

   a. Historical—collected before the period to be assessed (historical data can be good for baselines and to determine rates of change);
   b. Live—the data collections starts in the present and will be ongoing;
   c. Live with some historical data; or
   d. Complete—the entire period wanted is available and live data is collected?
   e. Could historical data be affected by changes to a platform's access rules or by differences in selection criteria over time?

5. How are datasets from multiple sources handled?

   a. Are the data from multiple sources analysed either together or using the same tools and queries.
   b. Are steps taken to avoid duplications that can inflate findings?
   c. Are the data selected using the same criteria for all sources?

The data base can affect information environment assessments both in obvious and subtle ways. Data collected to look at election disinformation would clearly hide a lot of disinformation related to COVID-19. On the other hand, a continuous data collection from all official Russian social media channels from 2010 onwards would facilitate analysis on the official spread of disinformation on any topic. How criteria are combined is also important. Data collected for the keywords "vaccine" **and** "COVID-19" could be used to find a range of conspiracy theories around the corona virus. If these keywords **and** the phrase "Bill Gates" is used, then results will mainly include conspiracy theories (or discussions thereof) that mentioned Bill Gates, a frequent target of conspiracy theories related to COVID-19 (Wakabayashi et al., 2020). Asking for "vaccine" **or** "COVID-19" **or** "Bill Gates" would result in a large dataset where the majority would have nothing to do with disinformation—it would include all mentions of Bill Gates in his role as Microsoft founder, any discussion on vaccines and all discussions on COVID-19.

8.8 Digging Deeper, Slowly

It is important to realise that data collection criteria will change over time; tactics and topics used for disinformation will frequently change. Detecting and analysing disinformation is an iterative process. One should therefore not expect to get all questions defined at the beginning. This in turn requires flexibility in the data collection. As the criteria changes it is important to re-check the data base to see if new questions can be answered through existing data selection criteria, or if one needs to have separate data collection for the new issues being investigated.

Using the COVID-19 crisis as an example, the core questions would have developed from a tentative, probing beginning, to more specific, health related issues over time:

- What is being said about COVID-19?
- Is anything negative being said about the official responses to COVID-19?
- What narratives are used?
- How widely spread is disinformation about the ability to handle the crisis?
- What groups are negative about vaccines and how large are these groups?
- Where do these posts come from?

8.9 Conclusion

The matters raised in this chapter will be well-known to experts in social media/big data analytics. However, issues that affect the relevance of datasets to information environment assessment are not necessarily discussed as a matter of course by data suppliers. Thus, there is a need to continuously examine what the data base is when presented with findings or offered a service.

This chapter suggests that despite the image of big data as a source of hard facts, when it comes to social media, it is rather soft and fluffy. Big data is not the same as "big social data" which can have many subtle nuances. Big data tools can too often give believable but incorrect answers due to the large quantities of such data and the inevitable large amount of noise—irrelevant data that is sucked into the sample, cluttering or even polluting the relevant data so that the graphics and automated analysis are less precise.

Data from social media have certain characteristics that distinguish them from more traditional datasets in terms of quantity (vast), the issues in a person's life they can cover (many) and the nature of the data (live and relational). Certain inherent technical, legal and commercial obstacles can affect the data collection in ways that reduces the data's relevance for a given task, in particular through data selection criteria that are not clearly communicated to customers. This necessitates an evaluation of the data base in order to determine how relevant a dataset is for the task at hand, typically an information environment assessment. It is proposed that such an evaluation can be done by first determining actual data requirements for an information environment assessment.

Evaluating the relevance of datasets for disinformation research is not necessarily a binary choice. Sometimes a less relevant dataset can be used to zero in on issues to pursue. For example, looking at available datasets on a related topic could help to discover narratives, hashtags and topic groups that might be vulnerable to disinformation. Anti-vaccination datasets are available from many academics who have studied this related disinformation phenomena, which could provide a better understanding of narratives, tactics, and groups that might be enlisted to promote disinformation about health issues including COVID-19. This can help to crystallise criteria required to get a more relevant dataset. A more relevant dataset can more readily be used to find answers to questions that arise in an information environment assessment. Such datasets are often freely available online and can also be useful to provide additional context and for historical comparisons.

The above checklist was developed in order to handle both the type of explicit selection criteria that disinformation research requires and the implicit selection choices that emerge from social media platforms' and data suppliers' business models and rules. The suggestion is that this list can be used to validate the relevance of datasets relative to the information environment assessment task at hand. It is hoped this will help researchers and analysts who must rely on data collected by others to better weigh the capabilities of third party data services to provide high quality data for deeper analysis.

References

Auxier, B., & Anderson, M. (2021, April 7). Social media use in 2021. *Pew Research Center: Internet, Science & Tech.* https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Bailey, J. (2019, September 12). Ninth circuit rules web scraping is not hacking. *Plagiarism Today*. https://www.plagiarismtoday.com/2019/09/12/ninth-circuit-rules-web-scraping-is-not-hacking/

Bay, S., & Biteniece, N. (2019). *The current digital arena and its risks to serving military personnel* (p. 14).

Berg, S. S., & Frivold, S. (2017). *Delingsøkonomiens virkning på eksisterndenæringer – Effekten av Airbnb på hotell-og serveringsvirksomhet i Norge*. Master's Thesis, Norges teknisk-naturvitenskapelige universitet. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2455674/Masteroppgave.Berg%26Frivold.pdf?sequence=1&isAllowed=y

Bergh, A. (2019). *Social network centric warfare: Understanding influence operations in social media* (FFI-Rapport No. 19/01194; p. 65). Norwegian Defence Research Establishment (FFI). http://hdl.handle.net/20.500.12242/2623

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports, 10*(1), 16598. https://doi.org/10.1038/s41598-020-73510-5

CrowdTangle. (2021). *Academics & researchers FAQ.* http://help.crowdtangle.com/en/articles/3323105-academics-researchers-faq

EU. (2018, May 25). *General data protection regulation (GDPR).* GDPR.Eu. https://gdpr.eu/tag/gdpr/

Greene, T., Shmueli, G., Ray, S., & Fell, J. (2019). Adjusting to the GDPR: The impact on data scientists and behavioral researchers. *Big Data*, *7*(3), 140–162. https://doi.org/10/ggdc43

Hageskal, A., & Ueland, T. H. (2018). *Nettpiratene, metoderapport—Dataskup 2018*. Skup.No. https://www.skup.no/sites/default/files/metoderapport/2018-10/Nettpiratene%2C%20Dagbladet.pdf

Hamilton, C., & Ohlberg, M. (2020). *Hidden hand: Exposing how the Chinese Communist Party is reshaping the world*. Simon and Schuster.

Ipsos. (2020). *Ipsos SOME Tracker* (Q4'19). Ipsos. https://www.ipsos.com/sites/default/files/ct/news/documents/2020-01/ipsos_some_4._kvartal_2019.pdf

Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A.-H. M., Hasan, S. M. M., Kabir, A., Yeasmin, D., Islam, M. A., Chowdhury, K. I. A., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19–Related infodemic and its impact on public health: A global social media analysis.

The American Journal of Tropical Medicine and Hygiene, *103*(4), tpmd200812. https://doi.org/10.4269/ajtmh.20-0812

Johnes, G. (2012, May). Harvesting tweets can open up a new world of valuable qualitative data. *British Politics and Policy at LSE*. http://blogs.lse.ac.uk/politicsandpolicy/good-uni-quality-nightlife-how-harvesting-tweets-opens-up-a-new-world-of-valuable-qualitative-data/

Kim, Y., Nordgren, R., & Emery, S. (2020). The story of goldilocks and three Twitter's APIs: A pilot study on Twitter data sources and disclosure. *International Journal of Environmental Research and Public Health*, *17*(3). https://doi.org/10.3390/ijerph17030864

Maldoff, G. (2016, April 19). How GDPR changes the rules for research. *The Privacy Advisor*. https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. *ArXiv:1306.5204 [Physics]*. http://arxiv.org/abs/1306.5204

Prisjakt. (2019). *Hvordan prisjakt sammenligner priser for deg*. Prisjakt. https://www.prisjakt.no/info/hvordan-vi-sammenligner-priser--i1

Richtel, M. (2020, February 6). W.H.O. fights a pandemic besides coronavirus: An 'infodemic'. *The New York Times*. https://www.nytimes.com/2020/02/06/health/coronavirus-misinformation-social-media.html

Wakabayashi, D., Alba, D., & Tracy, M. (2020, April 17). Bill gates, at odds with Trump on virus, becomes a right-wing target. *The New York Times*. https://www.nytimes.com/2020/04/17/technology/bill-gates-virus-conspiracy-theories.html