Vanderbilt Journal of Transnational Law

Volume 55
Issue 5 *November 2022*

Article 4

11-2022

Information Operations under International Law

Tsvetelina van Benthem
Oxford Institute for Ethics, Law and Armed Conflict

Talita Dias
University of Oxford

Duncan B. Hollis
Beasley School of Law, Temple Univ.

Information Operations under International Law

Tsvetelina van Benthem,* Talita Dias,** and Duncan B. Hollis***

ABSTRACT

An information operation or activity (IO) can be defined as the deployment of digital resources for cognitive purposes to change or reinforce attitudes or behaviors of the targeted audience in ways that align with the authors' interests. While not a new phenomenon, these operations have become increasingly prominent and pervasive in today's digital age, a trend that the ongoing war in Ukraine and the use of the internet for terrorist purposes tragically demonstrate. Against this backdrop, this Article critically assesses the existing international legal framework applicable to IOs. It makes three overarching claims. First, IOs can cause real and tangible harms to individual and state interests protected by international law. To prevent and remedy such harms, a robust and comprehensive legal framework constraining the use of IOs by both state and non-state actors becomes a necessity. Second, existing international law regulates IOs through a system of prohibitions, permissions, and requirements. In particular, the Article analyzes the extent to which international human rights law, the principles of non-intervention and sovereignty, and due diligence obligations apply to state and non-state uses of IOs. Third, the fact that existing international law captures some of the harms of IOs does not mean that this framework is sufficient or adequate. In fact, we argue that, in their current form, international rules on IOs are only partially effective given challenges relating to their (i) application, (ii) orientation, (iii) complexity, and (iv) enforcement in the context of information and communications technologies. While accepting that international law, both conventional and customary, already contains important protections against harmful IOs, our analysis aims to reignite a much-needed discussion of the merits and shortcomings that adopting a new regime tailored to IOs might produce.

---

* Tsvetelina van Benthem is a lecturer in Public International Law at the Oxford Diplomatic Studies Programme, and a researcher at the Oxford Institute for Ethics, Law and Armed Conflict (ELAC) and Merton College.

** Talita Dias is the Shaw Foundation Junior Research Fellow in Law at Jesus College, University of Oxford, and a Research Fellow at ELAC.

*** Duncan B. Hollis is Laura H. Carnell Professor of Law and Faculty Co-Director of the Institute for Law Innovation and Technology (iLIT) at Temple University's Beasley School of Law as well as a non-resident Fellow at the Carnegie Endowment for International Peace.

TABLE OF CONTENTS

I. INTRODUCTION

Information operations and activities (IOs) have become increasingly prominent and pervasive tools of power in today's digital age.[1] Spread through digital platforms, today's IOs can shape

1. We define an information operation as the deployment of digital resources for cognitive purposes to change or reinforce attitudes or behaviors of the targeted audience in ways that align with the authors' interests. *See* Duncan B. Hollis, *The Influence of War, The War for Influence*, 32 TEMPLE INT'L & COMP. L.J. 30, 35–36 (2018); *The Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities*, OXFORD INST. FOR ETHICS, L. AND ARMED CONFLICT pmbl. ¶ 3 (Feb. 7, 2022), https://www.elac.ox.ac.uk/the-oxford-process/the-statements-overview/the-oxford-statement-on-the-regulation-of-information-operations-and-activities/ [https://perma.cc/7FDH-7EK3] (archived Nov. 8, 2022) (defining an IO as

perceptions and influence interactions; they can inform and misinform; and they can bring together or isolate their target audiences. Their full import is on display in the recent—and alarming—examples of operations conducted via information and communications technologies (ICTs) prior to and throughout Russia's aggressive war in Ukraine.

The Russian government has sponsored IOs that seek to sow discord, demoralize the Ukrainian public, shatter Western alliances, and spread false information regarding the policies of the Ukrainian government and the conduct of the war.[2] Their contents range from unfounded claims of genocide against ethnic Russians by Ukrainian forces in the Donbas region,[3] to false allegations that Ukraine is controlled by "Nazis,"[4] and claims that atrocities committed by Russian forces in Ukrainian towns were staged.[5] Russia's campaigns have disseminated such information via both traditional and digital media, garnering the support of Russian policy makers, soldiers, and citizens for the war.[6] Despite evidence that these influence operations, when pursued with patience and persistence, are "well positioned" to exploit

---

"any coordinated or individual deployment of digital resources for cognitive purposes to change or reinforce attitudes or behaviours of the targeted audience"). For other definitional efforts, see KRISTINE BERZINA & ETIENNE SOULA, ALL. FOR SECURING DEMOCRACY, CONCEPTUALIZING FOREIGN INTERFERENCE IN EUROPE 6–7 (Mar. 18, 2020), https://securingdemocracy.gmfus.org/wp-content/uploads/2020/03/Conceptualizing-Foreign-Interference-in-Europe.pdf [https://perma.cc/W48B-45M3] (archived Oct. 15, 2022) (analyzing the definition of interference, along with the importance of having a definition for it, as it relates to cyber-attacks and social media); Barrie Sander, *Democracy Under the Influence: Paradigms of State Responsibility for Cyber Influence Operations on Elections*, 18 CHINESE J. INT'L L. 1, 11–15 (2019).

2. *See* Alden Wahlstrom, Alice Revelli, Sam Riddell, David Mainor & Ryan Serabian, *The IO Offensive: Information Operations Surrounding the Russian Invasion of Ukraine*, MANDIANT (May 19, 2022), https://www.mandiant.com/resources/information-operations-surrounding-ukraine [https://perma.cc/U6HT-S7GY] (archived Aug. 15, 2022).

3. *See* BBC Reality Check Team, *Ukraine Crisis: Vladimir Putin Address Fact-Checked*, BBC NEWS (Feb. 22, 2022), https://www.bbc.co.uk/news/60477712 [https://perma.cc/499F-ABZG] (archived Aug. 15, 2022).

4. *See* Alexey Kovalev, *Russia's Ukraine Propaganda Has Turned Fully Genocidal*, FOREIGN POL'Y (Apr. 9, 2022), https://foreignpolicy.com/2022/04/09/russia-putin-propaganda-ukraine-war-crimes-atrocities/ [https://perma.cc/968U-S96V] (archived Aug. 15, 2022); Mariia Kravchenko, *What Should Russia do with Ukraine?*, MEDIUM (Apr. 4, 2022), https://medium.com/@kravchenko_mm/what-should-russia-do-with-ukraine-translation-of-a-propaganda-article-by-a-russian-journalist-a3e92e3cb64 [https://perma.cc/H7NU-RYBE] (archived Aug. 15, 2022) (translation of a propaganda article by a Russian publication).

5. *See* Jeanne Whalen, Robyn Dixon & Mary Ilyushina, *Russia Denies and Deflects in Reaction to Bucha Atrocities*, WASH. POST (Apr. 4, 2022), https://www.washingtonpost.com/world/2022/04/04/russia-bucha-atrocities-war-crimes/ [https://perma.cc/N3
7M-EYU8] (archived Aug. 15, 2022).

6. *See* Kovalev, *supra* note 4.

existing polarizations,[7] their success among Ukrainian and Western audiences has so far been limited. This is likely due to the swift measures taken by the Ukrainian government and its allies to control social media narratives, expose atrocities, and build stories of their own modern-day heroes.[8] Ukrainians have launched their own information campaigns in turn to pierce the Kremlin's information curtain and reach the Russian population. Some of these have generated controversies—for example, calls for the death of—or otherwise hateful rhetoric targeting—Russian soldiers, leaders and even civilians on social media platforms;[9] online dissemination of videos depicting Russian POWs by both government and private accounts;[10] and the release of personal information of dead Russian soldiers.[11] The Russian invasion of Ukraine has made it abundantly clear how a war for dominant narratives can be as fierce and critical as the one fought on the ground.

States are not, however, the only actors tempted to unleash the power of modern IOs. Terrorist groups have been quick to take advantage of the digital environment as well. For more than two decades, states and scholars have waxed anxiously about the catastrophic prospects of "cyberterrorism" in the form of large-scale,

---

7. Brad Smith, *Defending Ukraine: Early Lessons from the Cyber War*, MICROSOFT ON THE ISSUES (June 22, 2022), https://blogs.microsoft.com/on-the-issues/2022/06/22/defending-ukraine-early-lessons-from-the-cyber-war/ [https://perma.cc/ZKG4-7RXK] (archived Aug. 15, 2022).

8. Consider, for example, the story of Ukrainian soldiers guarding Zmiinyi Island, who reportedly told a Russian warship to "go to hell" before they were presumably killed. *See Snake Island: Ukraine Says Troops who Swore at Russian Warship are Alive*, BBC NEWS (Feb. 28, 2022), https://www.bbc.co.uk/news/world-europe-60554959 [https://perma.cc/M33U-YTF4] (archived Aug. 15, 2022)

9. *See* Richard Lawler, *Facebook Allows Posts with Violent Speech Toward Russian Soldiers in Specific Countries*, VERGE (Mar. 10, 2022), https://www.theverge.com/2022/3/10/22970705/russia-ukraine-moderation-facebook-instagram-hate-speech-violence-policy [https://perma.cc/368A-29D5] (archived Aug. 15, 2022); Munsif Vengattil & Elizabeth Culliford, *Facebook and Instagram Let Users Call for Death to Russian Soldiers over Ukraine*, REUTERS (Mar. 10, 2022), https://www.reuters.com/world/europe/exclusive-facebook-instagram-temporarily-allow-calls-violence-against-russians-2022-03-10/ [https://perma.cc/38FY-9JVX] (archived Aug. 15, 2022); Dina Newman, *Facebook Struggles with Hate Speech Around Russia's Invasion of Ukraine*, MEDIA DIVERSITY INST. (Apr. 8, 2022), https://www.media-diversity.org/facebook-struggles-with-hate-speech-around-russias-invasion-of-ukraine/ [https://perma.cc/
8LXF-G82X] (archived Aug. 15, 2022).

10. *See* Aaron Blake, *Why You Should Think Twice before Sharing that Viral Video of an Apparent Russian POW*, WASH. POST (Mar. 7, 2022), https://www.washingtonpost.com/politics/2022/03/07/russian-pow-videos/ [https://perma.cc/CQ5G-W9XW] (archived Aug. 15, 2022).

11. *See* Laura Italiano, *Ukraine is Using Facial Recognition to ID Dead Russian Soldiers and Send Photos of Corpses Home to Their Moms: Report*, BUS. INSIDER (Apr. 15, 2022), https://www.businessinsider.com/ukraine-sending-photos-of-dead-russian-soldiers-home-moms-2022-4?r=US&IR=T [https://perma.cc/J3L5-8MXP] (archived Aug. 15, 2022).

one-off violent operations.[12] Fortunately, fears of terrorist groups' capacities to use online means to crash planes, derail trains, or poison water supplies have not yet come to fruition.[13] That said, transnational terrorist networks *have* repeatedly employed social media and online communication tools to incite violence, distribute propaganda, recruit, and finance their activities.[14] Labeled "the use of the internet to *facilitate* (but not perpetrate) terrorism,"[15] these acts can clearly be classified within the spectrum of modern IOs as well.

---

12. For early treatments, see John F. Murphy, *Computer Network Attacks by Terrorists: Some Legal Dimensions*, 76 INT'L L. STUD. 323, 325–26 (2002) (defining international terrorism and analyzing its definition, the laws that apply, and the steps to take to combat it); Dorothy Denning, *Activism, Hacktivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy*, *in* NETWORKS AND NETWARS: THE FUTURE OF TERROR, CRIME, AND MILITANCY 239, 281 (John Arquilla & David Ronfeldt eds., 2001) (analyzing how cyberterrorists use the internet and what influence they have been able to exert on policymakers); JOHN ROLLINS & CLAY WILSON, CONG. RSCH. SERV., RL33123, TERRORIST CAPABILITIES FOR CYBERATTACK: OVERVIEW AND POLICY ISSUES 3 (2007) (analyzing potential terrorist objectives and computer vulnerabilities, as well as terrorists' capabilities that could result in harm to the U.S.).

13. *See* Murphy, *supra* note 12, at 326–27; Ben Saul & Kathleen Heath, *Cyber Terrorism and Use of the Internet for Terrorist Purposes*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE 205, 225 (Nicholas Tsagourias & Russell Buchan eds., 2021); Heather A. Harrison Dinniss, *The Threat of Cyber Terrorism and What International Law Should (Try to) Do about It?*, 19 GEO. J. INT'L AFFS. 43–44, 47 (2018) (highlighting prospects that terrorists may use online means to cause blackouts or trigger lethal explosions but acknowledging "the most pressing cyber incidents are carried out by State actors" and highlighting Russian attribution of the TV5Monde operation that originally claimed to have terrorist sponsors).

14. *See* Saul & Heath, *supra* note 13, at 207–9; Scot A. Terban, *An Assessment of Violent Extremist Use of Social Media Technologies*, REAL CLEAR DEF. (Feb. 5, 2018) https://www.realcleardefense.com/articles/2018/02/05/an_assessment_of_violent_extremist_use_of_social_media_technologies_113015.html [https://perma.cc/9MZ4-DSMR] (archived Aug. 16, 2022) (detailing previous terrorists' use of social media and recommending a potential solution); Imran Awan, *Cyber-Extremism: Isis and the Power of Social Media*, 54 SOC'Y 138, 139–41 (2017); TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER-OPERATIONS 199 (Michael Schmitt ed., 2017) (describing the application of international law to cyber conflicts and cyber warfare) [hereinafter TALLINN MANUAL 2.0]; CATHERINE A. THEOHARY & JOHN ROLLINS, CONG. RSCH. SERV., RL41674, TERRORIST USE OF THE INTERNET: INFORMATION OPERATIONS IN CYBERSPACE 2–5 (2011).

15. Saul & Heath, *supra* note 13, at 207; *see also* U.N. OFF. ON DRUGS AND CRIME (UNODC), THE USE OF THE INTERNET FOR TERRORIST PURPOSES (Sept. 2012), https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf [https://perma.cc/QW4J-XML3] (archived Sept. 20, 2022); S.C. Res. 2129, ¶ 14 (Dec. 17, 2013) ("[noting] the evolving nexus between terrorism and information and communications technologies, in particular the Internet, and the use of such technologies to commit terrorist acts, and to facilitate such acts through their use to incite, recruit, fund, or plan terrorist acts"); G.A. Res. 60/288, § II, ¶ 12(a) (Sept. 20, 2006) (stating the UN's goal to "[c]oordinate efforts at the international and regional levels to counter terrorism in all its forms and manifestations on the Internet"); S.C. Res. 1373, ¶¶ 1, 2(e), 5 (Sept. 28, 2001) (laying out the UN's goal of punishing those who provide funding to terrorist organizations).

The diversity and effects of IOs may be most pronounced in wars and terrorist threats. But information campaigns can be initiated or facilitated by a much broader range of individuals, groups, corporations, and states for various ends—benign and malign—under disparate guises. The "infodemic" around COVID-19 vaccines dramatically impacted vaccine uptake amidst claims of microchips and risks of disease or death.[16] Such claims, like those on the inefficacy of masks or incentivizing the consumption of certain "miraculous" cures, risked significant harm to the life and health of individuals.[17] False claims may have equally significant (and harmful) outcomes in other areas—from manipulating electorates during democratic processes in order to favor particular positions or deny their actual outcome,[18] to altering perceptions of climate change or technological developments.[19] At its most severe, extremist content—and the appeal of broadcasting it online—has incentivized violence from Christchurch to Pittsburgh in what has become an all-too-common feature of modern life.[20]

---

16. *See, e.g.*, Lorna Christie, *Covid-19 Vaccine Misinformation*, UK PARLIAMENT POST (Apr. 26, 2021), https://post.parliament.uk/covid-19-vaccine-misinformation/ [https://perma.cc/QA4N-4WUN] (archived Aug. 16, 2022); *Fighting Misinformation in the Time of COVID-19, One Click at a Time*, WORLD HEALTH ORG. (Apr. 27, 2021), https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time [https://perma.cc/7CQQ-RXTF] (archived Aug. 16, 2022); Jack Goodman & Flora Carmichael, *Covid vaccine: 'Disappearing' needles and other rumours debunked*, BBC NEWS (Dec. 20, 2020), https://www.bbc.com/news/55364865 [https://perma.cc/3NQU-E89E] (archived Aug. 16, 2022).

17. *See, e.g.*, Nick Robins-Early, *Desperation, Misinformation: How the Ivermectin Craze Spread Across the World*, GUARDIAN (Sept. 24, 2021), https://www.theguardian.com/world/2021/sep/24/ivermectin-covid-peru-misinformation [https://perma.cc/28
6J-W5CQ] (archived Aug. 16, 2022); Rick Rouan, *Fact Check: Study Falsely Claiming Face Masks are Harmful, Ineffective is Not Linked to Stanford*, USA TODAY (Apr. 24, 2021), https://www.usatoday.com/story/news/factcheck/2021/04/24/fact-check-study-falsely-claiming-masks-harmful-isnt-stanfords/7353629002/ [https://perma.cc/APU2-N6JU] (archived Aug. 16, 2022).

18. *See generally* DEFENDING DEMOCRACIES: COMBATTING FOREIGN ELECTION INTERFERENCE IN A DIGITAL AGE (Duncan B. Hollis & Jens D. Ohlin eds., 2021); Sam Levine, *How Republicans Came to Embrace the Big Lie of a Stolen Election*, GUARDIAN (Sept. 13, 2021), https://www.theguardian.com/us-news/2021/jun/13/republicans-big-lie-us-election-trump [https://perma.cc/Z97B-CMW4] (archived Aug. 16, 2022).

19. *See, e.g.*, Kari Paul, *Climate Misinformation on Facebook 'Increasing Substantially', Study Says*, GUARDIAN (Nov. 4, 2021), https://www.theguardian.com/technology/2021/nov/04/climate-misinformation-on-facebook-increasing-substantially-study-says? [https://perma.cc/H35T-98KX] (archived Aug. 16, 2022).

20. *See, e.g.*, Jane Coaston, *The New Zealand Shooter's Manifesto Shows how White Nationalist Rhetoric Spreads*, VOX (Mar. 18, 2019), https://www.vox.com/identities/2019/3/15/18267163/new-zealand-shooting-christchurch-white-nationalism-racism-language [https://perma.cc/TS8C-FM3Y] (archived Aug. 16, 2022); Kevin Roose, *On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full*, N.Y. TIMES (Oct. 28, 2018), https://www.nytimes.com/2018/10/28/us/gab-robert-bowers-pittsburgh-synagogue-

As part of this symposium issue on the law of cyberterrorism, our Article explores international law's regulation of IOs including, but not limited to, those with terrorist origins. In doing so, we make three claims. First, international law currently regulates many IOs via prohibitions, permissions, and requirements. The principles of non-intervention and sovereignty conjoin with the obligation to respect human rights to constrain states from pursuing certain IOs themselves both at home and abroad. At the same time, duties to protect human rights and different rules featuring due diligence standards require states to take positive steps to respond to IOs emanating from their territory and jurisdiction. In other contexts (e.g., restrictions on freedom to receive and impart information), international law permits states to regulate IOs, but only subject to satisfying various conditions. We are now witnessing, moreover, a significant increase in state, intergovernmental, academic, and corporate initiatives to clarify *how* existing rules of international law apply to ICTs.[21] It is clear that the extant rules, with the benefit of cyber-specific interpretations, offer a comprehensive regulatory framework to address extremist content online as well as information campaigns that threaten things like public health or electoral processes.

Second, viewing the harm typically associated with terrorist activity through the lens of generally applicable international law rules and standards is necessary for very pragmatic reasons. To the extent international law regulates terrorist activity specifically, it has done so through piece-meal regulation of specific violent terrorist acts[22] (e.g., terrorist bombings[23] and seizures of aircrafts[24]), methods (e.g., the use

---

shootings.html [https://perma.cc/YUX2-AD9Z] (archived Aug. 16, 2022); Irene Khan (Special Rapporteur for Freedom of Opinion and Expression), *Disinformation and Freedom of Opinion and Expression - Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. Doc. A/HRC/47/25, ¶ 19 (Apr. 13, 2021) [hereafter Special Rapporteur Report on Disinformation and Freedom of Opinion].

21. *See, e.g.*, G.A. Res. 75/240 (Jan. 4, 2021); Rep. of the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of Int'l Sec., U.N. Doc. A/76/135, at 6 (2021); Rep. of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of Int'l Sec., U.N. Doc. A/70/174 (2015); TALLINN MANUAL 2.0, *supra* note 14; *The Oxford Statement on International Law Protections in Cyberspace*, *supra* note 1.

22. *See* Saul & Heath, *supra* note 13, at 212–17.

23. International Convention for the Suppression of Terrorist Bombings, Dec. 15, 1997, 2149 U.N.T.S. 256.

24. Convention for the Suppression of Unlawful Seizure of Aircraft, Dec. 16, 1970, 860 U.N.T.S. 105.

of plastic explosives[25] and nuclear material[26]), and financing.[27] Terrorism as a concept has, however, proven subjective, elusive, and divisive; it has effectively escaped a legally binding definition.[28] Elements such as the intention or purpose to compel become easily politicized, leaving negotiations at a dead end. As such, conventional and customary international law have no *general* definition of terrorism that we can apply in relevant online contexts, whether for delimiting internationally wrongful acts of states or invoking international or domestic crimes.[29] This, in turn, causes definitional difficulties when it comes to acts of *cyber*terrorism—that is, "the deliberate exploitation of computer networks as a means to launch a [terrorist] attack"[30]—or acts facilitative thereof, such as online terrorist propaganda or recruitment.[31] Only a few non-binding documents, most notably the Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online, have addressed the contours of online terrorist threats.[32]

Given the extant regulation of terrorism (and its focus on kinetic acts and methods), we posit that, at least in the short term, the

---

25. Convention on the Marking of Plastic Explosives for the Purpose of Detection, Mar. 1, 1991, 2122 U.N.T.S. 359.

26. Convention on the Physical Protection of Nuclear Material, Mar. 3, 1980, 1456 U.N.T.S. 124.

27. International Convention for the Suppression of the Financing of Terrorism, Dec. 9, 1999, 2178 U.N.T.S. 197.

28. At the international level, States have never agreed to a common definition for terrorism, let alone for cyberterrorism. *See* Dinniss, *supra* note 13, at 44. For its part, the United States has defined terrorism in terms of the employment of violent or dangerous acts. *See* Exec. Order No. 13,244, 66 Fed. Reg. 49079 (Sept. 23, 2001) (defining terrorism as "an activity that (1) involves a violent act or an act dangerous to human life, property, or infrastructure; and (2) appears to be intended to intimidate or coerce a civilian population; to influence the policy of a government by intimidation or coercion; or to affect the conduct of a government by mass destruction, assassination, kidnapping, or hostage-taking."). Online speech acts may not qualify as "violent" or "dangerous" even if they could trigger such behavior. At the same time, other (broader) definitions might treat online information and propaganda campaigns by transnational terrorist organizations as "cyberterrorism." *See* CATHERINE A. THEOHARY & JOHN ROLLINS, CONG. RSCH. SERV., R43955, CYBERWARFARE AND CYBERTERRORISM: IN BRIEF 1 (2015) (defining cyberterrorism as "the premeditated use of disruptive activities or the threat thereof, against computers and/or networks, with the intention to cause harm or further social, ideological, religious, political, or similar objectives, or to intimidate any person in furtherance of such objectives").

29. *See, e.g.*, Ben Saul, *Legislating from a Radical Hague: The United Nations Special Tribunal for Lebanon Invents an International Crime of Transnational Terrorism*, 24 LEIDEN J. INT'L L. 677 (2011) (analyzing the Appeals Chamber of the UN Special Tribunal of Lebanon use and interpretation of a customary international crime of transnational terrorism).

30. USE OF THE INTERNET FOR TERRORIST PURPOSES, *supra* note 15, at 11.

31. *See* Saul & Heath, *supra* note 13, at 205–10.

32. CHRISTCHURCH CALL, THE CHRISTCHURCH CALL TO ACTION TO ELIMINATE TERRORIST AND VIOLENT EXTREMIST CONTENT ONLINE (May 15, 2019), https://www.christchurchcall.com/christchurch-call.pdf [https://perma.cc/7ZC2-FL3J] (archived Aug. 18, 2022).

existing, general international legal framework applicable to IOs and other cyber activities will be a better vehicle for addressing the current online threat landscape occupied by transnational and domestic terror groups. Talking about conduct proscribed under generally applicable international law allows the international community to constrain harmful behavior while avoiding political stalemates surrounding use of the "terrorism" label or language.

Third, just because international law has extensive rules that address IOs does not mean that they are sufficient. These rules are, at best, only partially effective. In their current form, international rules on IOs face no less than four discrete challenges in the ICT environment: (i) application, (ii) orientation, (iii) complexity, and (iv) enforcement. Like all law, international law's *application* depends on facts. But, when it comes to IOs, the facts are often hard to ascertain. In particular, the widespread nature of such operations makes them difficult to identify in the first place. Once identified, attributing responsibility, whether technically or legally, presents particular challenges in cyberspace. At the same time, the quintessentially cognitive methods employed by IOs (i.e., to change or reinforce behavior) make it especially difficult to establish a causal link between operations and the harmful outcomes towards which they aim or ultimately lead.

But even if international law could surmount its operational challenges, its *orientation* around state behavior complicates its ability to regulate the IOs and activities of non-state actors, operations that form the lion-share of this problem set. International law covers some of this behavior via positive obligations, such as states' duties to protect and ensure human rights against violations by third parties. But these positive obligations are addressed to states, applying to non-state actors only indirectly. Furthermore, the law is often weighted to emphasize its prohibitions rather than its positive requirements;[33] and here those prohibitions only speak to states (or non-state actors acting under the instructions, direction, or control of states). As such, when it comes to both terrorist behavior and the role of social media platforms in hosting or amplifying harmful content, the law's state-centric orientation risks leaving out some of the major drivers of online harm.

Ironically, even as the law is insufficient in its application and orientation, it is exceedingly *complex*. As our review of the existing law will show, a single information operation may implicate multiple international legal rules and regimes to say nothing of domestic laws. Ambiguity often lurks at the edges of rules, and sometimes at their

---

33. *See* Duncan B. Hollis, *Re-Thinking the Boundaries of Law in Cyberspace: A Duty to Hack?*, *in* CYBERWAR: LAW & ETHICS FOR VIRTUAL CONFLICTS 154, 154–55 (Jens David Ohlin et al. eds., 2015).

very core. And even if the relevant rules can be specified with some degree of certainty, this does not solve the conundrum of interaction between specific rules and regimes. It is difficult for international lawyers to process all these issues, let alone on the short timelines that the risks posed by IOsmay require.

On top of all this, international law's *enforcement* mechanisms—the limited remedies available to respond to internationally wrongful acts—have yet to demonstrate a compliance pull to limit whether and when IOs occur or to facilitate means to remediate them. Simply put, as states and other stakeholders debate how international law regulates IOs, they ought to ask how well it does so. International lawyers can—and should—consider whether (and, if so, what) opportunities exist for more precise articulations of the law and its standards and even perhaps its progressive development. We recognize that there are real risks alongside any rewards that would accompany devising a *lex specialis* for IOs in the coming years. Yet we believe the rapid rise and importance of IOs to modern international relations warrants making such calculations and seeking collective solutions accordingly.

Our Article proceeds in four parts. In Part II, we define and explore the concept of an information operation and distinguish it from other information-related activities online. Part III examines some of the key, applicable rules of international law: positive and negative obligations under international human rights law, the principle of non-intervention, sovereignty, and two rules of general international law containing a due diligence standard—the *Corfu Channel* and the no-harm principles. Part IV identifies and elaborates the four challenges to the effectiveness of existing international law: application, orientation, complexity, and consequences. In Part V, we explore the potential benefits (and costs) of further clarifications or development of international law. We conclude with some thoughts on where and how such developments might occur.

II. WHAT ARE INFORMATION OPERATIONS AND WHY ARE THEY OF CONCERN?

While IOs have existed for centuries, they have garnered increasing attention over the last decade as states and other stakeholders have come to recognize the extent to which digital technologies facilitate their formation and execution. As yet, there is no internationally accepted definition of IOs (although it is no longer treated as including cyber-attacks or cyber surveillance operations, i.e.,

operations targeting computer systems or data).[34] For the purposes of this paper, we define IOs as the deployment of digital resources for cognitive purposes to change or reinforce attitudes or behaviors of the targeted audience in ways that align with the authors' interests.[35] Successful IOs do not necessarily threaten or intimidate targets. They influence, persuade, convince, or otherwise drive members of the targeted audience to adopt the goals that the IO author wishes them to adopt, whether by open or deceptive means.[36]

Under this definition, it becomes apparent that IOs are a regular feature of human relations. Families and friends regularly deploy online resources to get us to adopt or change our views, social norms, or political beliefs. Companies expend significant resources on marketing to convince us to buy their products and services. States deploy diplomacy, speeches, and other forms of strategic communication (including propaganda) to affect the behavior of adversaries and foreign populations.

The risks, however, are also apparent. Given the range of potential cognitive impacts IOs can generate, it becomes easy to see how they may threaten or result in a range of significant, real-world harms. IOs may destabilize electoral outcomes (e.g., falsehoods about the 2020 US presidential election catalyzing the right-wing occupation of the US Capitol on January 6, 2021).[37] They may undermine public health (e.g., the "infodemic" that has disrupted the "coordinated, medically sound response that is necessary to control the spread of the [COVID-19] virus").[38] They may even incite discrimination, violence, genocide, and other atrocities. Witness the dissemination of inaccurate and hateful rhetoric on Facebook against the Rohingya in Myanmar

---

34. In 2006, for example, the US Department of Defense defined information operations as those seeking to "influence, disrupt, corrupt, or usurp adversarial human and automated decision making while protecting [our] own," a definition that included all "computer network operations" including "computer network attacks." *See* JOINT CHIEFS OF STAFF, U.S. DEP'T OF DEF., JOINT PUB. 3-13, INFORMATION OPERATIONS, at ix (2006).

35. *See* Hollis, *The Influence of War*, *supra* note 1, at 35–36; *The Oxford Statement on International Law Protections in Cyberspace*, *supra* note 1, pmbl. ¶ 3.

36. *See* Herbert Lin & Jackie Kerr, *On Cyber-Enabled Information/Influence Warfare and Manipulation*, *in* THE OXFORD HANDBOOK ON CYBERSECURITY 251, 254 (Paul Cornish ed., 2021).

37. *See, e.g.*, Stuart A. Thompson, *Election Falsehoods Surged on Podcasts Before Capitol Riots, Researchers Find*, N.Y. TIMES (Jan. 4, 2022), https://www.nytimes.com/2022/01/04/technology/apple-google-spotify-podcast-election-misinformation.html [https://perma.cc/L859-YEQ5] (archived Sept. 20, 2022).

38. Marko Milanović & Michael N. Schmitt, *Cyber Attacks and Cyber (Mis)information Operations during a Pandemic*, 11 J. NAT'L SEC. L. & POL'Y 247, 249 (2020).

since 2017.[39] In 1994, broadcasts by Radio Télévision des Mille Collines radio in Rwanda would tell Hutus: "You have missed some of the enemies. You must go back there and finish them off. The graves are not yet full!"—a directive that would likely occur in an online forum if made today.[40]

What, then, are the types of IOs that can lead to such harms? There are multiple typologies. Three widely used categories differentiate IOs and other forms of "information disorder" based on the authors' intentions and the verifiability of the information deployed:

> (1) *Misinformation* – when false information is shared, but no harm is intended to arise from the sharing;
>
> (2) *Disinformation* – when false information is knowingly shared to cause harm; and
>
> (3) *Malinformation* – when verifiable information, personal views or opinions are shared to cause harm, including by moving information designed to stay private into the public sphere (e.g., doxing).[41]

Other ways of categorizing IOs focus on transparency—is the IO author's identity publicly known, anonymous, or affirmatively misrepresented? Misrepresented IO authors may create conditions for greater harms where audiences are more likely to be persuaded (or react) based on the assumed identity than if the author's true identity were known to them. Anonymous IO authors may also be problematic in some cases. Yet, it is important to note a long-standing tradition protecting anonymous speech (in the United States, such speech protection dates back to the framers of the US Constitution).[42] Finally,

---

39. *See* Human Rights Council, Rep. of the Independent International Fact-Finding Mission on Myanmar, ¶¶ 73–74, 84–89, U.N. Doc. A/HRC/39/64 (Sept. 12, 2018); *In Myanmar, "Pervasive Hate Speech and Shrinking Freedom"*, AL JAZEERA (Mar. 5, 2019), https://www.aljazeera.com/news/2019/3/5/in-myanmar-pervasive-hate-speech-and-shrinking-freedom [https://perma.cc/Q2CJ-R9DH] (archived Aug. 18, 2022); Steve Stecklow, *Why Facebook is Losing the War on Hate Speech in Myanmar*, REUTERS (Aug. 15, 2018), https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ [https://perma.cc/25T6-S2FA] (archived Aug. 18, 2022).

40. BILL BERKELEY, THE GRAVES ARE NOT YET FULL: RACE, TRIBE AND POWER IN THE HEART OF AFRICA 20 (2001).

41. CLAIRE WARDLE & HOSSEIN DERAKHSHAN, COUNCIL OF EUR., INFORMATION DISORDER: TOWARD AN INTERDISCIPLINARY FRAMEWORK FOR RESEARCH AND POLICYMAKING 5 (Sept. 27, 2017), https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77 [https://perma.cc/QS3P-PKJM] (archived Aug. 20, 2022); *see also* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 10–12, 15; EUROPEAN COMMISSION, EUROPEAN UNION, CODE OF PRACTICE ON DISINFORMATION (Sept. 2018).

42. For example, *The Federalist Papers* were signed under the same pseudonym—Publius—even though we now know they were authored by Alexander Hamilton, John Jay, and James Madison. *See generally* THE FEDERALIST.

some efforts look to isolate out particular types of speech based on its contents, most notably different forms of "hate speech," which are criminalized or otherwise prohibited across different legal systems.[43]

Whatever their form, in international relations, IOs employ cognitive methods to induce action or inaction in the target audience to further the author's political, economic, social, or cultural aims, including destabilizing the target.[44] They do this by taking advantage of various known cognitive and emotional biases that can be leveraged to influence individuals, leaders, groups, or networks. For example, human beings have a confirmation bias where we seek and interpret information in ways consistent with our existing attitudes and decisions, leading us to steer away from or discount inconsistent information.[45] We also share a loss-aversion bias—we act more recklessly to recoup perceived losses than in efforts to attain gains. As such, if we believe conditions are bad or deteriorating, we can be primed to act more recklessly.[46] Likewise, we are drawn to and tend to accept patterns, connections, or simple stories that we can clearly make sense of, but which are easily subject to manipulation or misinterpretation.[47] Alongside these cognitive biases, emotional biases can affect human reasoning and judgments. We are often emotionally uncomfortable when there are inconsistencies between our behavior and our beliefs. As such, humans often avoid information that challenges their beliefs, seek out behavior that bolsters their beliefs, or rationalize their behavior to be consistent with their beliefs.[48]

---

43. *See, e.g.*, Luvell Anderson & Michael Barnes, *Hate Speech*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., 2022) (analyzing how to define hate speech, the impact of hate speech, and what steps can be taken to combat it).

44. *See* Lin & Kerr, *supra* note 36, at 254–55.

45. *See generally* Kate Sweeny, Darya Melnyk, Wendi Miller & James Shepperd, *Information Avoidance: Who, What, When, and Why*, 14 REV. GEN. PSYCH. 340 (2010); Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCH. 175 (1998).

46. *See generally* Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSPS. 193 (1991); Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979).

47. *See The Mind, Explained: Brainwashing* (Netflix Nov. 19, 2021).

48. *See* Lin & Kerr, *supra* note 36, at 256; *see also* William Hart, Dolores Albarracín, Alice H. Eagly, Inge Brechan, Matthew J. Lindberg & Lisa Merrill, *Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information*, 135 PSYCH. BULL. 555, 555–56 (2009) (analyzing whether exposure to information is bolstered by defense or accuracy motives); Sweeny, Melnyk, Miller & Shepperd, *supra* note 45, at 340–53. At the same time, the science is not perfect; conclusions drawn about biases from some long-touted experiments have proven hard to replicate, while some individuals or portions of a population may be more resistant to influence than others. *See, e.g.*, Open Science Collaboration, *Estimating the Reproducibility of Psychological Science*, 349 SCI. 943, 947 (2015) (presenting data to show the importance of reproducibility and that no one single indicator describes replication success); Lin &

In sum, although IOs are a regular—and often valuable—form of human interaction, the cognitive behavioral effects they can generate create strategic opportunities for those looking to cause harm to different persons or objects.

III. HOW DOES INTERNATIONAL LAW APPLY TO IOS?

Far from an absence of international law in the information space, there are a plethora of extant obligations derived from both conventional and customary international law. This Part briefly surveys the five most prominent sources of legal obligations relevant to IOs: (i) negative and positive duties under international human rights law, (ii) the principle of non-intervention, (iii) sovereignty and due diligence standards found in (iv) the *Corfu Channel* and (v) no-harm principles.[49]

It is important to emphasize at the outset that *all* of these international legal obligations require attribution of certain wrongful conduct to a state. Attribution is the process of identifying who is responsible for the activity or conduct in question, including the lack of care in preventing, stopping, or redressing a certain harm. In the online environment, attribution has discrete technical, political, and legal forms.[50] Technical attribution can involve identifying (i) the machine from which a cyber-operation originates, (ii) the operator of that machine, or (iii) the person or entity who directed it.[51] Political

---

Kerr, *supra* note 36, at 255 ("[T]here will always be people in a target population that are immune to its effects—this is most true in populations that have strong institutions and traditions dedicated to the rule of law and relatively sane trustworthy (i.e., not corrupt) political leaders.").

49. Although comprehensive, our treatment is not exhaustive. Other international law rules, such as the prohibition on the use of force or even the right of self-defense in response to an armed attack, might be relevant in certain circumstances. We believe such circumstances will arise, however, quite rarely, if at all, given existing questions about whether and what effects online, especially those arising from IOs, might qualify as uses of force. *See* Milanović & Schmitt, *supra* note 38, at 258–61. Similarly, although IOs relating to elections may implicate the right of self-determination, we do not address that law in detail here. *See generally* Jens David Ohlin, *Did Russian Cyber Interference in the 2016 Election Violate International Law?*, 95 TEX. L. REV. 1569 (2017).

50. *See* Martha Finnemore & Duncan B. Hollis, *Beyond Naming and Shaming? Accusations and International Law in Cybersecurity*, 33 EURO. J. INT'L L. 969, 976 (2021).

51. DAVID WHEELER & GREGORY LARSEN, INST. FOR DEF. ANALYSES, TECHNIQUES FOR CYBER ATTACK ATTRIBUTION (Oct. 2003), https://apps.dtic.mil/sti/pdfs/ADA468859.pdf [https://perma.cc/PX3D-XJ2P] (archived Sept. 20, 2022); Kristen E. Eichensehr, *The Law & Politics of Cyberattack Attribution*, 67 UCLA L. REV. 520, 528 (2020).

attribution refers to the decision-making processes by which an attribution is exposed in public or diplomatic settings.[52]

For international law purposes, however, the question is one of legal attribution—assigning legal responsibility to the activity in question. A wide variety of actors lie behind harmful IOs. Individual hackers, groups of hackers, armed groups with an online presence, and state agencies all can—and do—engage in IOs that fall within the scope of international rules. However, the rules reviewed here—obligations under human rights law, the principles of non-intervention and sovereignty, and the *Corfu Channel* and no-harm principles—all require a breach, whether by an act or omission, legally attributable to a state to qualify as an internationally wrongful act.

When is a state legally responsible for IOs or related activities? States are always responsible for operations conducted by their own *de jure* organs (e.g., intelligence agencies or military forces). States are also responsible for the failure of their own official organs to exercise due diligence in preventing, stopping, or redressing IOs that meet the requisite threshold of harm under relevant primary obligations.[53] IOs by non-state actors may also be legally attributed to a state in at least two circumstances. First, non-state actors that are not part of the official state hierarchy can still qualify as *de facto* organs if they stand in a position of "complete dependence" in relation to a state. According to the International Court of Justice (ICJ), "complete dependence" denotes a lack of "any real autonomy,"[54] the group being "merely the instrument" of the state.[55] Importantly, once the qualification as a state organ (be it *de jure* or *de facto*) is made, even conduct where non-state actors exceed the state's grant of authority or contravene its instructions will still be attributed to the state.[56] Second, even when it is impossible to characterize a certain non-state group as an organ of a

---

52. *See* Thomas Rid & Ben Buchanan, *Attributing Cyber-Attacks*, 38 J. STRATEGIC STUD. 4, 4 (2014); JASON HEALEY, ATL. COUNCIL, BEYOND ATTRIBUTION: SEEKING NATIONAL RESPONSIBILITY FOR CYBER ATTACKS 3–7 (Jan. 2012), https://www.atlanticcouncil.org/wp-content/uploads/2012/02/022212_ACUS_NatlResponsibilityCyber.PDF [https://perma.cc/75ZS-X624] (archived Sept. 20, 2022).

53. *See* Antonio Coco & Talita de Souza Dias, *'Cyber Due Diligence': A Patchwork of Protective Obligations in International Law*, 32 EURO. J. INT'L L. 771, 771–72, 777–78 (2021).

54. Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosn. & Herz./Serb. & Montenegro), Merits, 2007 I.C.J. Rep. 243, ¶¶ 393–94 (Feb. 26).

55. *Id.* at ¶ 394.

56. *See* G.A. Res. 56/83, Responsibility of States for Internationally Wrongful Acts, art. 7 (Dec. 12, 2000) [hereinafter ARSIWA].

state, particular conduct of that group can still be attributed to the state if made under its direction, instructions, or control.[57]

In sum, applying existing international law to IOs will not depend on who authors the IO so much as on the state's behavior vis-à-vis the specific operation or its perpetrators.

A. *International Human Rights Law: Negative and Positive Obligations*

We begin with the regime protecting the interests of individuals—international human rights law—for three reasons. First, since IOs by definition operate at the individual level (i.e., they target individuals *qua* individuals or as agents/representatives of states, firms, and other institutional actors) it makes sense to examine existing laws designed to ensure respect for, and protection of, rights at the individual level.

Second, although the UN General Assembly adopted a consensus resolution in 2014 recognizing that "the same rights that people have offline must also be protected online,"[58] states and scholars have given relatively little attention to the application of these rules to cyberspace generally, let alone IOs specifically.[59]

Third, international human rights law provides fertile ground for assessing the human impact of IOs, as well as the obligations of states to uphold those rights in conducting IOs and protecting individuals from IOs carried out by others. Given the pervasiveness of the internet and other ICTs across the globe—to say nothing of our dependence on them—IOs may affect a wide array of human rights both inside and outside a state's territory, including several individual freedoms as well as social, economic, and cultural rights.[60]

What behaviors can individuals expect from states pursuant to these human rights obligations? Broadly speaking, a recognized human right entails two types of duties. First, it entails negative duties for states (i.e., duties not to engage in a particular form of conduct). Second, it creates positive duties for states (i.e., duties to take certain steps to safeguard rights-holders from harm, including when such

---

57. *See id.* art. 8. In addition, state responsibility may arise where a non-state actor exercises elements of government authority, acts in the absence or default of official authorities, or engages in conduct that is acknowledged and adopted by a state as its own. *Id.* arts. 5, 9, 11.

58. G.A. Res. 68/167, The Right to Privacy in the Digital Age, ¶ 3 (Jan. 21, 2014).

59. For example, only five of the 154 rules detailed in TALLINN MANUAL 2.0 focus on international human rights. *See* TALLINN MANUAL 2.0, *supra* note 14, 179–208 (Rules 34–38).

60. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 22.

harm originates from non-state actors, as well as to create the conditions for the enjoyment of all rights).[61]

1. Negative Human Rights Obligations

Many IOs originate from state authorities who direct them against the state's own population. Recent examples abound. Witness the downplaying of the transmissibility and lethality of COVID-19[62] in the United States,[63] Brazil,[64] and Nicaragua.[65] Other domestic-oriented IOs involve debunking climate change,[66] official incitement or endorsement of violence or discrimination,[67] and unsubstantiated allegations of election fraud.[68] When made by state authorities such

---

61. *See* Hum. Rts. Comm., General Comment No. 31 [80], The Nature of the General Legal Obligation Imposed on States Parties to the Covenant, U.N. Doc. CCPR/C/21/Rev.1/Add., ¶¶ 6–8 (May 26, 2004).

62. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 49.

63. *See* Juana Summers, *Timeline: How Trump Has Downplayed The Coronavirus Pandemic*, NPR (Oct. 2, 2020), https://www.npr.org/sections/latest-updates-trump-covid-19-results/2020/10/02/919432383/how-trump-has-downplayed-the-coronavirus-pandemic [https://perma.cc/7CAW-KD9C] (archived Aug. 23,2022).

64. *See How Bolsonaro Downplayed Covid-19 Before, and After, He Contracted the Virus*, GUARDIAN (July 8, 2020), https://www.theguardian.com/world/video/2020/jul/08/how-bolsonaro-downplayed-covid-19-before-and-after-he-contracted-the-virus-video [https://perma.cc/GFG7-K626] (archived Aug. 23, 2022).

65. *See* Wilfredo Miranda, *Sandinista Leaders Fall Victim to Coronavirus Outbreak They Downplayed*, GUARDIAN (June 8, 2020), https://www.theguardian.com/world/2020/jun/08/nicaragua-coronavirus-sandinista-leaders-fall-victim [https://perma.cc/MC5F-VW4D] (archived Aug. 23, 2022).

66. *See* ROYAL SOC'Y, THE ONLINE INFORMATION ENVIRONMENT - UNDERSTANDING HOW THE INTERNET SHAPES PEOPLE'S ENGAGEMENT WITH SCIENTIFIC INFORMATION 7, 9, 32, 88 (Jan. 2022), https://royalsociety.org/-/media/policy/projects/online-information-environment/the-online-information-environment.pdf [https://perma.cc/YLZ9-ADBL] (archived Sept. 20, 2022); *Facebook's Climate of Deception: How Viral Misinformation Fuels the Climate Emergency*, AVAAZ (May 11, 2021), https://secure.avaaz.org/campaign/en/facebook_climate_misinformation/ [https://perma.cc/LVX4-KYP3] (archived Aug. 23, 2022); Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 25;*"I Don't Think Science Knows": Trump Denies Climate Change Link to Wildfires – Video'*, GUARDIAN (Sept. 15, 2020), https://www.theguardian.com/us-news/video/2020/sep/15/i-dont-think-science-knows-trump-denies-climate-change-link-to-wildfires-video [https://perma.cc/M9F3-FKHG] (archived Aug. 23, 2022).

67. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 26, 48; Ryan Goodman, Mari Dugas & Nicholas Tonckens, *Incitement Timeline: Year of Trump's Actions Leading to the Attack on the Capitol*, JUST SEC. (Jan. 11, 2021), https://www.justsecurity.org/74138/incitement-timeline-year-of-trumps-actions-leading-to-the-attack-on-the-capitol/ [https://perma.cc/T7HL-W56B] (archived Aug. 23, 2022).

68. *See US Election 2020: Trump's Voting Fraud Claims Explained*, BBC NEWS, https://www.bbc.co.uk/news/av/world-us-canada-54835475 [https://perma.cc/CK6M-FPL9] (archived Aug. 23, 2022).

claims can have serious adverse consequences for the life and health of individuals, their trust in democratic institutions, and their right to free and fair elections.[69] This is because the greater the prominence, power, or influence of the speaker, the higher the likelihood that their target audience will act upon the claim.[70]

Other state-sponsored IOs target foreign audiences. Examples include claims, allegedly originating from China and Russia, that Western democracies are failing and that the West is too weak to respond to the pandemic.[71] As a response to British Broadcasting Corporation reports on alleged human rights violations against the Uyghurs in China, an influence operation linked to the Chinese Communist Party sought to discredit the broadcasting company.[72] In some cases, IOs may operate broadly for both domestic and international audiences. For example, while the Twenty-Sixth Climate Change Convention was under way, false claims about climate change spread widely—promoted by advertisers—on Facebook.[73]

International human rights law binds states to a range of *negative* obligations.[74] IOs may implicate such obligations, including those stemming from the rights to life, health, privacy, and freedom of thought; the right to seek and impart information; the right to freely participate in democratic processes; and the prohibition of ill-treatment.[75] These are detailed, *inter alia,* in the International

---

69. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 24, 49.

70. *See Dangerous Speech: A Practical Guide*, DANGEROUS SPEECH PROJECT, (Apr. 19, 2021), https://dangerousspeech.org/guide/ [https://perma.cc/PP6H-GJHN] (archived Aug. 23, 2022).

71. *See* Mark Scott, *Russia and China Push 'Fake News' Aimed at Weakening Europe: Report*, POLITICO (Apr. 1, 2020), https://www.politico.eu/article/russia-china-disinformation-coronavirus-covid19-facebook-google/ [https://perma.cc/ZT5V-ZBN3] (archived Aug. 23, 2022).

72. *See* Matt Burgess, *China Aims Its Propaganda Firehose at the BBC*, WIRED (Aug. 18, 2021), https://www.wired.com/story/china-aims-its-propaganda-firehose-at-the-bbc/ [https://perma.cc/Q6VM-BTDB] (archived Aug. 23, 2022).

73. *See* Elizabeth Culliford, *During COP26, Facebook Served Ads with Climate Falsehoods, Skepticism*, REUTERS (Nov. 18, 2021), https://www.reuters.com/business/cop/during-cop26-facebook-served-ads-with-climate-falsehoods-skepticism-2021-11-18/ [https://perma.cc/DKZ3-JW7G] (archived Aug. 23, 2022); Luke Hurst, *COP26: 'Staggering Scale' of Climate Misinformation on Facebook Revealed in New Report*, EURONEWS (Nov. 5, 2021) https://www.euronews.com/next/2021/11/05/cop26-staggering-scale-of-climate-misinformation-on-facebook-revealed-in-new-report [https://perma.cc/UGW3-4TGM] (archived Aug. 23, 2022).

74. *See* Dinah Shelton & Ariel Gould, *Positive and Negative Obligations*, *in* THE OXFORD HANDBOOK OF INTERNATIONAL HUMAN RIGHTS LAW 562, 563 (Oxford Univ. Press 2013). While reference will be made to specific human rights law instruments, the rights examined in this article are also protected under customary international law.

75. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 22, 24–25. For a detailed analysis of these and other rights, see Talita

Covenant on Civil and Political Rights,[76] in various regional human rights instruments,[77] and under customary international law.[78]

For starters, the right to life prohibits arbitrary deprivation of life.[79] Deprivation of life, according to the Human Rights Committee, "involves intentional or otherwise foreseeable and preventable life-terminating harm or injury, caused by an act or omission."[80] Depending on the circumstances—including its content, source, virality, and means of dissemination—information (and the conduct it may instigate or prevent) can cause as much harm as direct physical acts. That state agents may have, orally or in writing, incentivized a population to imbibe toxic detergents as a cure for a potentially lethal disease instead of causing life-threatening harm by beating individuals with batons should not be a relevant distinction for the purposes of negative state obligations.

Consider the January 6, 2021, events in Washington, DC. After a long campaign of eroding trust in democratic elections through an overwhelming wave of Twitter activity, the rhetoric of then-President Trump culminated in a riot that threatened and effectively interfered with the life and health of state officials and ordinary citizens. The events of January 6th were not a random occurrence materializing

---

de Souza Dias, Antonio Coco & Tsvetelina van Benthem, *Background Paper: The Oxford Covid-19 Vaccine (CHADOX1 NCOV-19) Development Stages and Applicable Protective Obligations under International Law*, 153, July 2020; *see also* Talita de Souza Dias & Tsvetelina van Benthem, *Background Paper: Online Electoral Disinformation: A Human Rights Law Perspective*, 251, Oct. 2020.

76. International Covenant on Civil and Political Rights art. 19, Dec. 16, 1966, 999 U.N.T.S. 171.

77. *See* African Charter of Human and Peoples' Rights arts. 8–9, Oct. 21, 1986, 1520 U.N.T.S. 217; American Convention on Human Rights art. 4, Nov. 22, 1969, 1144 U.N.T.S. 123; European Convention on Human Rights art. 10, Nov. 4, 1950, 213 U.N.T.S. 222.

78. *See, e.g.*, U.N. Hum. Rts. Comm., CCPR Gen. Comment No. 24(52): Issues Relating to Reservations Made upon Ratification or Accession to the Covenant or the Optional Protocols thereto, or in Relation to Declarations under Article 41 of the Covenant, U.N. Doc. CCPR/C/21/Rev.1/Add.6, ¶ 8 (Jan. 4, 1994) (suggesting that, *inter alia*, the rights to life, freedom of thought and the prohibition of incitement to violence, discrimination and hostility are part of customary international law) [hereinafter CCPR General Comment 24(52)].

79. *See* International Covenant on Civil and Political Rights, *supra* note 76, art. 6; African Charter of Human and Peoples' Rights, *supra* note 77, art. 4; American Convention on Human Rights, *supra* note 77, art. 4; European Convention on Human Rights, *supra* note 77, art. 2. Of note, the European Convention on Human Rights (ECHR) regulates deprivation of life through limited exceptions rather than an 'arbitrariness' standard.

80. Hum. Rts. Comm. General Comment No. 36 (2018) on article 6 of the International Covenant on Civil and Political Rights, on the right to life, U.N. Doc. CCPR/C/GC/36, ¶ 6 (Oct. 30, 2018) [hereinafter HRC General Comment 36].

through an unfortunate alignment of stars.[81] They were the outcome of a persistent and manipulative information operation coming from the Oval Office that instilled fear and hatred in Trump supporters.[82] At the sentencing hearing for one of the January 6th rioters, a US federal judge called the mob "pawns" that were "called to Washington, DC, by an elected official and prompted to walk to the Capitol by an elected official."[83] Here, the state itself endangered its own people. Indeed, there is a growing understanding that domestic terrorism is a real and tangible threat—one that should not be ignored, especially given its online drivers. In June 2021, the Biden administration issued its National Strategy for Countering Domestic Terrorism, which underscored the capacity of social media to amplify threats to public safety.[84]

The right to life has a particular relevance to a host of IOs with content ranging from climate change to medical disinformation to intercommunal animosity. Death and injury are not required to engage the rule's prohibitions; engaging in life-threatening behavior is sufficient. In interpreting the European Convention on Human Rights, for instance, the European Court of Human Rights (ECtHR) has made clear that the right to life can be engaged even if the applicant did not die—behavior that puts the applicant's life at serious risk also qualifies.[85] Likewise, in its General Comment No. 36, the Human Rights Committee concluded that "[t]he obligation of States parties to respect and ensure the right to life extends to reasonably *foreseeable* threats and life-threatening situations that *can* result in loss of life."[86]

---

81. *See* Ed Kilgore, *Trump's Long Campaign to Steal the Presidency: A Timeline – The Insurrection was a Complex, Yearslong Plot, not a One-Day Event. And It Isn't Over*, N.Y. MAG. (Feb. 3, 2022), https://nymag.com/intelligencer/article/trump-campaign-steal-presidency-timeline.html [https://perma.cc/TA9K-3MPY] (archived Aug. 24, 2022).

82. *See* Atlantic Council's DFRLab, *#StopTheSteal: Timeline of Social Media and Extremist Activities Leading to 1/6 Insurrection*, JUST SEC. (Feb. 10, 2021), https://www.justsecurity.org/74622/stopthesteal-timeline-of-social-media-and-extremist-activities-leading-to-1-6-insurrection/ [https://perma.cc/R322-NMBT] (archived Aug. 24, 2022).

83. Hannah Rabinowitz, *Federal Judge Says Trump has Responsibility for January 6, Calling Rioter a 'Pawn'*, CNN (Nov. 19, 2021), https://edition.cnn.com/2021/11/19/politics/judge-blames-trump-riot/index.html [https://perma.cc/R322-NMBT] (archived Aug. 24, 2022).

84. *See* NAT'L SEC. COUNCIL, STRATEGY FOR COUNTERING DOMESTIC TERRORISM 9 (June 2021), https://www.whitehouse.gov/wp-content/uploads/2021/06/National-Strategy-for-Countering-Domestic-Terrorism.pdf [https://perma.cc/7E5Y-9CMZ] (archived Sept. 20, 2022).

85. *See* Makaratzis v. Greece, App. No. 50385/99, ¶ 55 (Dec. 20, 2004), https://hudoc.echr.coe.int/eng?i=001-67820 [https://perma.cc/YXE7-EL55] (archived Aug. 24, 2022).

86. HRC General Comment 36, *supra* note 80, ¶ 7 (emphasis added); *see also Makaratzis*, App. No. 50385/99 ¶¶ 6, 17, 21–22, 63.

The right to health is equally relevant in the IO context. General Comment 14 of the Committee on Economic, Social and Cultural Rights recognizes that

> violations of the obligation to respect are those State actions, policies or laws that contravene the standards set out in article 12 of the Covenant and are *likely* to result in bodily harm, unnecessary morbidity and preventable mortality. Examples include . . . *the deliberate withholding or misrepresentation of information vital to health protection or treatment.*[87]

In the (different) context of electoral processes, state-distributed false information could interfere with other rights, such as the rights to freedom of thought and opinion as well the right to freedom of expression. Freedom of expression, moreover, includes the right to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, through any media. While the protection of this right is not limited to accurate or innocuous information,[88] the proliferation of misinformation, disinformation, and malinformation may interfere with individuals' ability to think and form opinions freely. It may also have a chilling effect on the public's willingness to seek, receive, and express different types of information online and offline.[89] This is particularly the case for online hate speech and doxing, which can lead to the silencing of dissenting voices, especially vulnerable groups such as racial or ethnic minorities, women, and members of the LGBTQ+ community.[90] In the same vein, continuous exposure to dis- and misinformation may affect an individual's right to be properly informed,[91] undermining one of the core normative foundations of free speech (i.e., to enable individuals to challenge established truths and foster the possibility of new truths).[92] To be sure, speech emanating from state actors as well as public figures, such as opposition parties,

---

87. Comm. On Econ., Soc. and Cultural Rts., General Comment No. 14: The Right to the Highest Attainable Standard of Health, U.N. Doc. E/C.12/2000/4, ¶ 50 (Aug. 11, 2000) (emphasis added) [hereinafter ESCR General Comment 14].

88. *See* Hum. Rts. Comm., General Comment No. 34, Article 19: Freedoms of Opinion and Expression, U.N. Doc. CCPR/C/GC/34, ¶¶ 11, 47–49 (July 29, 2011) [hereinafter HRC General Comment 34].

89. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 23–24, 27–28, 33–36, 66.

90. *See* Dubravka Šimonović, (Special Rapporteur on Violence against Women), *Rep. of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective*, ¶¶ 29, 73, U.N. Doc. A/HRC/38/47 (June 18, 2018); Irene Khan (Special Rapporteur for Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, ¶ 27, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018).

91. *See* The Sunday Times v. United Kingdom, 2 Eur. Ct. H.R. (ser. A) 245, 281 (1979).

92. *See, e.g.*, JOHN STUART MILL, ON LIBERTY 19–24 (Batoche Books, 2001).

may enjoy heighted protection as "political speech."[93] This is true insofar as the disclosure of the speech in question is of public interest to society. However, as the ECtHR has noted, even political speech is not absolute and may be limited in a necessary and proportionate manner to protect the rights and reputations of others, including from violence and defamation.[94]

As noted by the 2017 Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda, while the freedom of expression is

> not limited to "correct" statements, [but] also protects information and ideas that may shock, offend and disturb . . . this does not justify the dissemination of knowingly or recklessly false statements by official or State actors; . . .
>
> State actors should not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false (disinformation) or which demonstrate a reckless disregard for verifiable information (propaganda).
>
> State actors should, in accordance with their domestic and international legal obligations and their public duties, take care to ensure that they disseminate reliable and trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment.[95]

This suggests that states may breach their negative human rights obligations even if they do not know that the information they disseminate is indeed false, insofar as they are reckless or should have known about the inaccuracy of their claims.[96] In the same vein, malinformation could violate states' duties to respect human rights if its dissemination is reckless.

The right to non-discrimination on grounds such as race, nationality, religion, gender, or sexual orientation may also be affected

---

93. *See* Ceylan v. Turkey, App. No. 23556/94, 15 Eur. Ct. H.R. 1061 ¶ 34 (1999); Mouvement Raëlien Suisse v. Switzerland, App. No. 16354/06, 2012, Eur. Ct. H.R. ¶ 61 (July 13, 2012), https://hudoc.echr.coe.int/eng?i=001-112165 [https://perma.cc/5CVB-UMBW] (archived Aug. 24, 2022).

94. *See* Ivanović v. Montenegro, App. No. 24387/10, ¶¶ 61–66 (June 5, 2018); Pastörs v. Germany, App. No. 55225/14, 2019, ¶¶ 38–42 (Oct. 3, 2019), https://hudoc.echr.coe.int/eng?i=001-196148 [https://perma.cc/3FEM-47NF] (archived Aug. 24, 2022).

95. *Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda*, United Nations (U.N.) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, pmbl. ¶ 7, operative ¶¶ 2(c)–(d), (Mar. 3, 2017) [hereinafter Joint Declaration on Freedom of Expression and "Fake News," Disinformation and Propaganda].

96. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 88.

by IOs that are designed to spread hate. This is especially the case of online content that goes beyond mere advocacy for hatred to include incitement to discrimination, hostility, or violence,[97] which is often the case in the context of acts that facilitate terrorism.

2. Positive Human Rights Obligations

In the digital age, where access to ICTs is enabled by (and granted to) a wide range of private entities, it is evident that IOs impacting human rights are not exclusive to states. Non-state actors can just as easily, and more often, take advantage of the opportunities offered by online platforms (and, in particular, of the operation of ranking and recommendation algorithms) to mount, contribute to, or enable large-scale information campaigns. These can, in turn, lead to significant harm. Consider QAnon, a far-right conspiracy theory that garnered significant support in the United States and which reportedly sprang into existence without any foreign assistance. In relation to the COVID-19 pandemic, QAnon influencers on Twitter promoted the "Mineral Miracle Supplement," advertised as a product that can prevent COVID-19 symptoms and sold by the Texas-based Genesis II Church of Health and Healing.[98] The US Food and Drug Administration had previously issued a warning about the potentially life-threatening side effects of that supplement.[99]

Even though the source of such IOs may be a private entity rather than a state, states are bound under international human rights law to protect the rights whose enjoyment such operations may imperil. According to the Human Rights Committee, "the duty to protect life also implies that States parties should take appropriate measures to address the general conditions in society that may give rise to direct threats to life or prevent individuals from enjoying their right to life with dignity."[100] Similarly, under the right to health, states must "take

---

97. *See* U.N. Hum. Rts. Comm., Views adopted by the Committee under Article 5(4) of the Optional Protocol, concerning communication No. 2124/2011, U.N. Doc. CCPR/C/117/D/2124/2011, ¶ 10.4 (Mar. 29, 2017).

98. *See* Marc-André Argentino, *Qanon Conspiracy Theories about the Coronavirus Pandemic are a Public Health Threat*, THE CONVERSATION (Apr. 8, 2020), https://theconversation.com/qanon-conspiracy-theories-about-the-coronavirus-pandemic-are-a-public-health-threat-135515 [https://perma.cc/Y2MJ-ZTVW] (archived Aug. 15, 2022).

99. *See* FDA News Release, *FDA Warns Consumers about the Dangerous and Potentially Life Threatening Side Effects of Miracle Mineral Solution*, FOOD & DRUG ADMIN. (Aug. 12, 2019), https://www.fda.gov/news-events/press-announcements/fda-warns-consumers-about-dangerous-and-potentially-life-threatening-side-effects-miracle-mineral [https://perma.cc/NR58-9YWP] (archived Aug. 15, 2022).

100. HRC General Comment 36, *supra* note 80, ¶ 26. One aspect of these "general conditions" is the "prevalence of life-threatening diseases." *See also id.* ¶¶ 8, 62 (noting that states are required to provide "quality and evidence-based information and

all necessary measures to safeguard persons within their jurisdiction from infringements of the right to health by third parties."[101]

To protect and ensure the genuine and effective exercise of the right to receive and impart information, states must not only refrain from interfering with speech acts but also are required to take positive steps by law or action, serving as the ultimate guarantor of pluralism in society.[102] In the field of audiovisual broadcasting, the freedoms of expression and information require states to ensure public access to impartial and accurate information and a "range of opinion and comment."[103] States must also prohibit by law "any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground."[104] States are also required to implement special measures to protect individuals from racial discrimination, including its prohibition by law and "immediate and effective measures, particularly in the fields of teaching, education, culture and information, with a view to combating prejudices which lead to racial discrimination."[105]

There is no prescriptive list of the positive actions required, and they vary significantly according to each right and the particular context of application. States are not required to do the impossible or to discharge a "disproportionate burden,"[106] but they are expected to adopt measures that are available and reasonable under the circumstances.[107] Such flexibility, however, does not detract from all

---

education about sexual and reproductive health" and "environment hazards," respectively, to fulfil their positive duty to protect life). In its previous general comment on the right to life, the Committee noted that "the right to life has been too often narrowly interpreted. The expression 'inherent right to life' cannot properly be understood in a restrictive manner, and the protection of this right requires that States adopt positive measures. In this connection, the Committee considers that it would be desirable for States parties to take all possible measures to . . . adopt[] measures to eliminate . . . epidemics." U.N. Hum. Rts. Comm., CCPR Gen. Comment No. 6: Article 6 (Right to Life), ¶ 5 (Apr. 30, 1982), https://www.refworld.org/docid/45388400a.html [https://perma.cc/Z89S-HNBC] (archived Sept. 20, 2022).

101. ESCR General Comment 14, *supra* note 87, ¶ 51.

102. *See* Manole v. Moldova, App. No. 13936/02, Eur. Ct. H.R. ¶ 99 (2009); Özgür Gündem v. Turkey, App. No. 23144/93, Eur. Ct. H.R. ¶ 43 (2000).

103. *Manole*, App. No. 13936/02 ¶ 100.

104. International Covenant on Civil and Political Rights, *supra* note 76, art. 26; *see also* Hum. Rts. Comm., General Comment 18: Non-discrimination (Nov. 10, 1989), https://www.refworld.org/docid/453883fa8.html [https://perma.cc/W79U-T6Y2] (archived Sept. 20, 2022).

105. International Convention on the Elimination of All Forms of Racial Discrimination art. 7, Dec. 21, 1965, 660 U.N.T.S. 14.

106. Osman v. U.K., App. No. 87/1997/871/1083, Eur. Ct. H.R. ¶ 16 (1998); *see also* Tănase v. Romania, App. No. 41720/13, Eur. Ct. H.R. ¶ 136 (June 25, 2019), https://hudoc.echr.coe.int/eng?i=001-194307 [https://perma.cc/X6D8-2AXH] (archived Sept. 20, 2022).

107. *See* McCann v. U.K., App. No. 18984/91, Eur. Ct. H.R. ¶ 151 (1995); Rodriguez v. Honduras, Merits, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 4, ¶ 167 (July

states having an obligation to take necessary and feasible measures to prevent, mitigate, and redress harm arising from IOs that foreseeably impact the enjoyment of human rights.[108] It is irrelevant whether the operation is conducted by a non-state actor or another state, or even simply caused by an accident such as a natural or human disaster—the obligation to protect does not depend on the nature and characteristics of the source of harm. Rather, its trigger lies in the existence of a "reasonably foreseeable threat" to a specific right.[109] Since the obligation is breached by an omission, reasonable foreseeability, coupled with a state's capacity to take potentially effective measures, suffices to link the state's conduct to the IO's human rights impact.[110]

There are many human rights-related steps that a state can take to tackle the threat of harmful IOs. For instance, it can require online providers to implement mechanisms for combating or mitigating the spread of false claims, in line with the rights to freedom of thought, opinion, and expression. Recently, for example, Spotify decided to add content advisories to all podcasts dealing with COVID-19 and to direct listeners to trusted sources of information.[111] Even though this move was welcomed by the White House, then–Press Secretary Jen Psaki urged the company to do more.[112] In lieu of piecemeal approaches to what online platforms can and should do, states could enact clear and transparent legal frameworks demanding specific action to prevent,

---

29, 1988); *see also* LETTER FROM THE MINISTER OF FOREIGN AFFAIRS TO THE PRESIDENT OF THE HOUSE OF REPRESENTATIVES ON THE INTERNATIONAL LEGAL ORDER IN CYBERSPACE 4 (July 5, 2019), https://ceipfiles.s3.amazonaws.com/pdf/CyberNorms/LawStatements/Letter+to+the+Parliament+on+the+International+Legal+Order+in+Cyberspace+(Statement+by+the+Netherlands).pdf [https://perma.cc/9QFR-9M2E] (archived Sept. 20, 2022) [hereinafter Dutch Foreign Ministry Letter to Parliament]; REPUBLIC OF KOREA, COMMENTS ON THE PRE-DRAFT OF THE OEWG REPORT 5 (Apr. 14, 2020), https://ceipfiles.s3.amazonaws.com/pdf/CyberNorms/UNGGE/Republic+of+Korea+Comments+on+the+Pre-Draft+of+the+OEWG+Report.pdf [https://perma.cc/Y947-8PWP] (archived Sept. 20, 2022).

108. *See* Coco & de Souza Dias, *supra* note 53, at 796, 800.

109. HRC General Comment 36, *supra* note 80, ¶ 18; Coco & de Souza Dias, *supra* note 53, at 799.

110. *See* Vladislava Stoyanova, *Causation between State Omission and Harm within the Framework of Positive Obligations under the European Convention on Human Rights*, 18 HUM. RTS. L. REV. 309, 309–46 (2018); Vladyslav Lanovoy, *Causation in the Law of State Responsibility*, 90 BRITISH Y.B. INT'L L. 22–26 (2022).

111. *See* Ramishah Maruf, *Spotify Makes Public Platform Rules that Cover Covid-19 Misinformation. Will it be Enough?*, CNN BUSINESS (Jan. 31, 2022), https://edition.cnn.com/2022/01/30/business/spotify-rules-joe-rogan-reliable-sources/index.html [https://perma.cc/36C3-5PYC] (archived Aug. 19, 2022).

112. *See* John Bowden, *White House urges Spotify to Take Further Action on Joe Rogan: "More can be Done"*, INDEPENDENT (Feb. 2, 2022), https://www.independent.co.uk/news/world/americas/us-politics/joe-rogan-spotify-covid-white-house-b200548
8.html [https://perma.cc/FP3R-DRBQ] (archived Aug. 19, 2022).

stop, and redress harmful or illegal IOs consistently with international human rights law. In addition to requiring such specific steps from companies, states may consider more general, preventative approaches, such as strategies to build trust in government and science, as well as foster critical and resilient audiences—key determinants in curbing the spread of harmful IOs.[113] Prime examples of such measures are media and information literacy campaigns.

The relationship between IOs and states' positive obligations under international human rights law may be quite complex. This is because, as seen earlier, different human rights may require different protective measures with respect to different aspects of mis-, dis-, and malinformation. More importantly, as speech-based acts, IOs themselves may be protected by the right to freedom of expression under different human rights treaties and customary international law. At the same time, upholding freedom of expression, especially by promoting a free, independent, and plural media environment, can be a powerful way to counter harmful IOs.[114] Thus, any state action to protect human rights from IOs must carefully consider their impact on perpetrators' and audiences' rights to seek, receive, and impart information.[115] Different regulatory frameworks will apply depending on the disseminated content, with varying degrees of discretion left to the state.

a. IOs That Must Be Prohibited by States

International human rights law prohibits certain categories of speech, including in the context of IOs,[116] requiring states to enact domestic prohibitions.[117] For instance, Article 20 of the International Covenant on Civil and Political Rights (ICCPR) provides that

> 1. Any propaganda for war shall be prohibited by law.
>
> 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.[118]

---

113. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 94; G.A. Res. 75/267, ¶ 3 (Mar. 29, 2021).

114. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 4, 38, 93–94; G.A. Res. 75/267, *supra* note 113.

115. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 30.

116. *See id.* ¶ 43.

117. A prohibition is not, however, tantamount to criminalization, and criminalization should be reserved for the most serious of crimes. *See* U.N. Secretary-General, *Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 8, U.N. Doc. A/74/486 (Oct. 9, 2019) [hereinafter U.N. Secretary-General on Freedom of Opinion and Expression].

118. International Covenant on Civil and Political Rights, *supra* note 76, art. 20.

Article 20 is a manifestation of the right to non-discrimination and is regarded as *lex specialis* to the right to freedom of expression, laid out in ICCPR Article 19.[119] It *requires* states to *prohibit* by law (though not necessarily criminal law) propaganda for wars of aggression and incitement to discrimination, hostility, or violence on the basis of race, religion, or nationality. According to the UN High Commissioner for Human Rights and the UN Special Rapporteur on Freedom of Expression, the severity of the legal sanction (civil, criminal, or administrative) applied against the incitement must be necessary and proportionate to the seriousness of the act, taking into account the status of the speaker, their intent, the audience, the content and form of the speech, its reach, and the risk of harm.[120]

Similarly, Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination requires states to condemn all propaganda and all organizations that are based on ideas or theories of superiority of one race or group of persons of one color or ethnic origin, or that attempt to justify or promote racial hatred and discrimination in any form.[121] This includes an obligation to *criminalize* all dissemination of ideas based on *racial* superiority or hatred, incitement to racial discrimination, and all acts of violence or incitement to such acts against any race or group of persons of *another color or ethnic origin*, including the provision of any assistance to racist activities and their financing.[122]

As the text of those provisions makes clear, states have little or no discretion when it comes to sanctioning racially, religiously, or nationality-motivated incitement; legislative action is imperative.[123] Granted, some states, including Australia, Belgium, the United Kingdom, and the United States, have made reservations to both

---

119. *See* HRC General Comment 34, *supra* note 88, ¶ 51.

120. *See* U.N. High Commissioner for Human Rights, *Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred*, ¶¶ 20, 29, 34, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013) ("Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."); U.N. Secretary-General on Freedom of Opinion and Expression, *supra* note 117, ¶ 14; Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 43.

121. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 44.

122. *See* International Convention on the Elimination of All Forms of Racial Discrimination, *supra* note 105, art. 4(a).

123. *See* Hum. Rts. Comm., General Comment No. 11: Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred (Art. 20), ¶ 2, (July 19, 1983); HRC General Comment 34, *supra* note 88, ¶ 51; JEROEN TEMPERMAN, RELIGIOUS HATRED AND INTERNATIONAL LAW: THE PROHIBITION OF INCITEMENT TO VIOLENCE OR DISCRIMINATION 74–77 (James Crawford & John S. Bell eds., 2016); Rebecca Meyer, *Pursuing a Universal Threshold for Regulating Incitement to Discrimination, Hostility or Violence*, 44 BROOK. J. INT'L L. 310, 322 (2018).

paragraphs of Article 20.[124] These reserve the right of states to either (i) apply Article 20 in line with the right to freedom of expression under Article 19, (ii) not to introduce *further* legislation on the matter, or (iii) not to adopt *any* legislation or action that would conflict with domestic law such as the US Constitution.[125] As others have noted, none of these reservations reject the substance of Article 20 (i.e., that incitement to violence, discrimination, or hostility is unlawful and must be tackled).[126] While reservations of the first type are not in any way contrary to Article 20 (insofar as this provision must be applied in line with the requirements of Article 19(3)),[127] the latter two do carve out an important aspect of Article 20's implementation: the requirement to enact legislation prohibiting war propaganda and incitement.[128] Yet it is questionable whether reservations to Article 20 are effective, insofar as the provision is part of customary international law.[129] In any event, the domestic laws of the states that have made reservations to Article 20 do contemplate, to a greater or lesser extent, prohibitions of different forms of incitement.[130] The US Supreme Court has found, for example, that incitement laws are not contrary to the Constitution's First Amendment if they prohibit "imminent lawless action."[131]

Incitement is an inchoate intentional conduct. This means that while dis- and malinformation may amount to such prohibited acts, misinformation is, by definition, excluded, given the absence of an intention to cause harm. Along similar lines, the 2017 Joint Declaration on Freedom of Expression and "Fake News," Disinformation and Propaganda expresses concern

> that disinformation and propaganda are often designed and implemented so as to mislead a population, as well as to interfere with the public's right to know and the right of individuals to seek and receive, as well as to impart, information

---

124. *See* International Covenant on Civil & Political Rights, *supra* note 76; *see also* TEMPERMAN, *supra* note 123, at 72–73. Similar reservations have been made with respect to International Convention on the Elimination of All Forms of Racial Discrimination, *supra* note 105, art. 4. The reservation made by the United States to Article 4 reads: "That the Constitution and laws of the United States contain extensive protections of individual freedom of speech, expression and association. Accordingly, the United States does not accept any obligation under this Convention, in particular under articles 4 and 7, to restrict those rights, through the adoption of legislation or any other measures, to the extent that they are protected by the Constitution and laws of the United States." *See id.*

125. *See* TEMPERMAN, *supra* note 123, at 72–73.

126. *See id.* at 73.

127. *See* U.N. High Commissioner for Human Rights, *supra* note 120, ¶¶ 17–18; HRC General Comment 34, *supra* note 88, ¶ 48.

128. *See* TEMPERMAN, *supra* note 123, at 73; Meyer, *supra* note 123, at 323–24.

129. *See* CCPR General Comment 24(52), *supra* note 78, ¶ 8 (stating that "provisions in the Covenant that represent customary international law [...] may not be the subject of reservations," and listing the prohibition of "advocacy of national, racial or religious hatred" among customary provisions).

130. *See* TEMPERMAN, *supra* note 123, at 77–79.

131. Brandenburg v. Ohio, 395 U.S. 444, 449 (1969).

> and ideas of all kinds, regardless of frontiers, protected under international legal guarantees of the rights to freedom of expression and to hold opinions;
>
> [and emphasizes] that some forms of disinformation and propaganda may harm individual reputations and privacy, or incite to violence, discrimination or hostility against identifiable groups in society.[132]

At the same time, the declaration concludes that

> [c]onsideration should be given to protecting individuals against liability for merely redistributing or promoting, through intermediaries, content of which they are not the author and which they have not modified.[133]

An example of an IO amounting to prohibited content is the type of claim that spread with the pandemic (especially at its start) blaming certain ethnic or national groups for COVID-19, accompanied by incitement to violence towards members of these groups.[134] Similarly, certain types of online racist abuse targeting England's football players following the Euro 2020 Championship final clearly amounted to incitement to racial discrimination, hostility or violence, or expressions of racial superiority.[135] Moreover, in the context of terrorism and online facilitative acts, ISIS has used a range of mainstream social media platforms, such as Facebook, Twitter, YouTube, and TikTok (along with smaller websites), to directly call for, and justify, violence against different ethnic, religious, and national groups.[136]

---

132. Joint Declaration on Freedom of Expression and "Fake News," Disinformation and Propaganda, *supra* note 95, at 1.

133. *Id.* ¶ 1(e).

134. *See* Mark Townsend & Nosheen Iqbal, *Far Right Using Coronavirus as Excuse to Attack Asians, Say Police*, GUARDIAN (Aug. 29, 2020), https://www.theguardian.com/society/2020/aug/29/far-right-using-coronavirus-as-excuse-to-attack-chinese-and-south-east-asians [https://perma.cc/EQK3-AL4K] (archived Aug. 22, 2022).

135. *See* Talita de Souza Dias & Sahil Thapa, *Tackling Football-Related Online Hate Speech: The Role of International Human Rights Law: Part I*, EJIL: TALK! (July 30, 2021), https://www.ejiltalk.org/tackling-football-related-online-hate-speech-the-role-of-international-human-rights-law-part-i/ [https://perma.cc/7M79-G6WV] (archived Aug. 22, 2022); Talita de Souza Dias & Sahil Thapa, *Tackling Football-Related Online Hate Speech: The Role of International Human Rights Law: Part II*, EJIL: TALK! (July 30, 2021), https://www.ejiltalk.org/tackling-football-related-online-hate-speech-the-role-of-international-human-rights-law-part-ii/ [https://perma.cc/7A8C-U7H7] (archived Aug. 22, 2022).

136. *See* Awan, *supra* note 14, at 138–49; Brendan I. Koerner, *Why ISIS is Winning the Social Media War*, WIRED (Apr. 2016), https://www.wired.com/2016/03/isis-winning-social-media-war-heres-beat/ [https://perma.cc/X22V-DEKB] (archived Aug. 22, 2022); Saul & Heath, *supra* note 13, at 207, 209.

b. IOs That May Be Prohibited or Otherwise Limited by States

IOs that are not prohibited under international human rights law benefit from the regime's protection. As a starting point, states are bound to respect and protect freedom of expression. However, international law allows—but does not require—states to prohibit or otherwise limit speech acts that may harm or threaten other protected interests. For such speech limitations to be lawful and consistent with the right to freedom of expression, they must

> (i) be grounded in a sufficiently clear legal basis,
>
> (ii) pursue a legitimate aim,
>
> (iii) be necessary (in the sense of the least restrictive measure possible) as well as
>
> (iv) proportionate to the importance of the interest or right protected.[137]

Under ICCPR Article 19(3), the legitimate aims for limiting freedom of expression are (a) "respect of the rights or reputations of others" and (b) "the protection of national security or of public order (*ordre public*), or of public health or morals."[138] These requirements apply not only to the very definitions of prohibited or limited speech acts but also to the respective measures to tackle them.[139] This means that states must articulate in sufficiently clear and accessible laws what content may—or must—be limited by the state itself or online intermediaries, what legitimate purposes justify such limitations, and what measures are deemed necessary and proportionate to limit such speech acts.[140] These may range from severe measures such as content takedowns and user suspension to less serious ones, including labeling, de-prioritization,[141] or "digital nudges" such as fact-check alerts or content suggestions.[142] States cannot simply rely on platform terms of service or community

---

137. Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 39.

138. International Covenant on Civil & Political Rights, *supra* note 76, art. 19.

139. *See* HRC General Comment 34, *supra* note 88, ¶¶ 24, 27.

140. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 40–41; U.N. Secretary-General on Freedom of Opinion and Expression, *supra* note 117, ¶¶ 31–32.

141. *See* Ben Whitelaw, *Talita Dias on Tackling Hate Speech with Civil and Political Rights*, EVERYTHING IN MODERATION (Feb. 23, 2022), https://www.everythinginmoderation.co/talita-dias-civil-political-rights/ [https://perma.cc/V5U8-J7XL] (archived Aug. 22, 2022).

142. *See* Markus Weinmann, Christoph Schneider & Jan vom Brocke, *Digital Nudging*, 58 BUS. INFO. SYS. ENG'G 433, 433–35 (2016).

guidelines in meeting their obligations to respect and protect freedom of expression.[143]

IOs falling short of incitement to violence, hostility, or discrimination that may still be prohibited or limited for a legitimate purpose potentially include instances of disinformation, malinformation, and even misinformation—though, as seen earlier, the latter should, as a general rule, be exempt from liability.

c. Residual Forms of IOs That Must Be Protected by States

Under the ICCPR, all residual speech acts, including IOs, that *fall below* the thresholds of mandatory or optional prohibitions or limitations (regulated by ICCPR Articles 20 and 19(3), respectively) must be protected under ICCPR Article 19(2). After all, freedom of expression is the rule and limitations the exception. That said, there are *no* types of speech acts that fall under this category *per se* (and thus would be absolutely immune from limitation). Virtually any type of otherwise free speech act can be justifiably limited under ICCPR Article 19(3) if the conditions assessed earlier are met (i.e., legality, legitimacy, necessity, and proportionality).

Nevertheless, certain types of speech do receive heightened protection under the law in different contexts, as even minor restrictions may prove unnecessary and/or disproportionate. Such speech includes content critical of institutions or religious tenets, religious or political satire, and speech or content whose political importance or public interest outweighs other protected interests.[144] Of course, this does not mean that states cannot (or should not) seek to address the root causes of IOs implicating such subjects. This can be done by employing, *inter alia*, educational, digital, and media literacy strategies, awareness-raising campaigns, counter-speech tactics, competition laws or regulations, and/or advertisement policies, alongside robust public information regimes to empower individuals,

---

143. *See* Talita Dias, *Hate Speech and the Online Safety Bill: Ensuring Consistency with Core International Human Rights Instruments 9–15*, U.K. HOUSE OF COMMONS, DIGITAL, CULTURE, MEDIA AND SPORT SUB-COMMITTEE ON ONLINE HARMS AND DISINFORMATION (Sept. 2021), https://committees.parliament.uk/writtenevidence/38393/pdf/ [https://perma.cc/9WJU-N6XC] (archived Aug. 22, 2022); Talita Dias, *Amending Online Safety Bill to Ensure Consistency with Core International Human Rights Instruments: Specific Recommendations 6–8*, U.K. HOUSE OF COMMONS, DIGITAL, CULTURE, MEDIA AND SPORT SUB-COMMITTEE ON ONLINE HARMS AND DISINFORMATION (Oct. 2021), https://committees.parliament.uk/writtenevidence/39923/pdf/ [https://perma.cc/8SCS-HRVL] (archived Sept. 20, 2022).

144. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 42, 79; *Ceylan*, App. No. 23556/94 ¶ 34; *Mouvement Raëlien Suisse*, App. No. 16354/06 ¶ 61; Castells v. Spain, App. No. 11798/85, Eur. Ct. H.R. ¶ 43 (1992); Wingrove v. the United Kingdom, App. No. 17419/90, Eur. Ct. H.R. ¶ 58 (1995); *see also* Milanović & Schmitt, *supra* note 38, at 276–77.

foster diversity, and build trust and resilience in societies.[145] As seen earlier, many human rights, such as the right to non-discrimination, may in fact *require* such forms of preventative action that do not interfere with the right of freedom of expression.

Having all restrictive measures reviewed against the standards of legality, legitimacy, necessity, and proportionality mandated by international human rights law constitutes an important bulwark for the realization of these rights. The ICCPR, for instance, recognizes the remedial right of individuals. Specifically, Article 2(3) requires states to ensure an effective remedy for human rights violations, as determined by judicial, administrative, legislative, or other competent legal authorities, as well as to develop the possibilities of a *judicial* remedy. This means that individuals whose human rights have been infringed upon by IOs—as well as those whose freedom of expression has been violated in the process of curbing such operations—must be able to present their claims to competent state authorities, preferably judicial bodies. It is not enough for states to rely on platforms' automated content moderation systems, given their inability to identify the nuances of language and context that are key to distinguishing between different types of IOs.[146] It is also not enough for states to require online platforms to put in place complaint or appeal mechanisms against harmful content or wrongful content takedowns or limitations.[147] States themselves must make available official avenues for redress, whether these are judicial or out-of-court processes, such as mediation, conciliation, or arbitration.[148]

Regional human rights instruments have their own thresholds for limitations and safeguards to freedom of expression. In the context of the European Convention on Human Rights (ECHR), for example, Article 10(2) authorizes limitations to freedom of expression insofar as these are prescribed by law and necessary in a democratic society

---

145. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶¶ 21, 23; U.N. Secretary-General on Freedom of Opinion and Expression, *supra* note 117, ¶¶ 18, 24, 28, 54.

146. *See* Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 71.

147. *See* U.N. Secretary-General on Freedom of Opinion and Expression, *supra* note 117, ¶¶ 33, 57(e); U.N. High Commissioner for Human Rights, *supra* note 120, ¶¶ 33–34.

148. *See, e.g.*, *Germany: Network Enforcement Act Amended to Better Fight Online Hate Speech*, LIBR. OF CONG., https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-online-hate-speech/ (last visited Sept. 20, 2022) [https://perma.cc/MQE4-TJ9E] (archived Aug. 24, 2022); Netzdurchsetzunggesetz [NetzDG] [Network Enforcement Act], Sept. 1, 2017, BGBL. at 1182, § 3(5)–(9), https://germanlawarchive.iuscomp.org/?p=1245 [https://perma.cc/U3NG-3D8R] (archived Aug. 24, 2022); Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC of 15 Dec. 2020, annex COM(2020) 825 [hereinafter EU Digital Services Act].

> in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[149]

In cases involving hate speech, however, such as *Garaudy v. France* and *Pastörs v. Germany* (both on Holocaust denial), as well as *I.A. v. Turkey* (on "an abusive attack on the Prophet of Islam"),[150] the ECtHR seems to have resorted to the ECHR's "abuse of rights" clause in Article 17, rather than assessing whether limitations to freedom of expression are lawful under Article 10. In doing so, the Court suggests that certain types of "abusive" content fall outside the protective scope of the right to freedom of expression, effectively sidestepping the four-part balancing test (i.e., legality, legitimacy, necessity, and proportionality) required under Article 10.[151]

When it comes to mis- and disinformation, the ECtHR referred explicitly to "fake news" in *Brzeziński v. Poland*.[152] It did so in the context of local elections in Poland and a statement made by a candidate for a local government position towards the outgoing local administration. In particular, the Court considered Poland's election law which allows a court, within twenty-four hours, to consider whether certain published information qualifies as "untrue" and, if so, to issue an order prohibiting its further distribution.[153] While a violation was found on the basis of the procedure before the Polish courts and the sanction imposed, the ECtHR also recognized the necessity of combatting the dissemination of false information on electoral candidates in view of preserving the integrity of the public debate.[154]

Similarly, a recent Council of Europe Recommendation notes that "disinformation undermines trust in the media and threatens the

---

149. European Convention for the Protection of Human Rights and Fundamental Freedoms art. 10(2), Nov. 4, 1950, 213 U.N.T.S. 222.

150. *See* Garaudy v. France, App. No. 65831/01, 2003-IX, Eur. Ct. H.R. ¶¶ 23–24; Pastörs v. Germany, App. No. 55225/14, 2019-V, Eur. Ct. H.R. ¶¶ 36–49; I.A. v. Turkey, App. No. 42571/98, 2005-VIII, Eur. Ct. H.R. ¶¶ 29–32.

151. *See* David Keane, *The Innocence of Satirists: Will Caricatures of the Prophet Mohammad Change the ECHR Approach to Hate Speech?*, EJIL:TALK! (Sept. 26, 2012), https://www.ejiltalk.org/the-innocence-of-satirists-will-caricatures-of-the-prophet-mohammad-change-the-echr-approach-to-hate-speech/ [https://perma.cc/3F5K-YDB7] (archived Aug. 24, 2022).

152. Brzeziński v. Poland, App. No. 47542/07, Eur. Ct. H.R. ¶¶ 28–29, 35, https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-194958%22]} [https://perma.cc/TX9E-MART] (archived Sept. 20, 2022).

153. *Id.* ¶ 55.

154. *See id.* ¶ 55.

reliability of information that feeds public debate and democracy."[155] Thus, it concludes that "[c]oncerted national and/or transnational efforts to address disinformation and propaganda should receive full support from States in a manner that does not undermine their independence."[156] Dissenting Opinions of ECtHR judges in *Benitez Moriana and Iñigo Fernandez v. Spain* and *Rashkin v. Russia* go as far as to suggest that, to be protected, information must be true and critical assertions must have some factual basis.[157]

For its part, the European Union (EU) draws a line between "illegal content" and false claims that are not necessarily illegal. Under European Commission Recommendation 2018/334 of 1 March 2018 (on measures to effectively tackle illegal content online) (the Recommendation), examples of illegal content include child pornography and terrorist propaganda. But the Recommendation's definition of "illegal content" is otherwise quite broad: "[A]ny information which is not in compliance with Union law or the law of a Member State concerned."[158] Part of the Recommendation's covered content does overlap with content that the ICCPR requires states to prohibit.[159] But the wide scope of "illegal content" could potentially also include IOs that are prohibited under domestic law but do not amount to content that must or may be prohibited under the ICCPR or the ECHR, such as misinformation. For illegal content under the recommendation, the EU has outlined a notice and counter-notice procedure for the assessment of content by hosting providers.[160]

Building on this existing framework, the EU Digital Services Act requires content takedowns only with respect to illegal content, following a notice-and-action process.[161] But, again, illegal content

---

155. COUNCIL OF EUR., RECOMMENDATION CM/REC(2022)4 OF THE COMMITTEE OF MINISTERS TO MEMBER STATES ON PROMOTING A FAVOURABLE ENVIRONMENT FOR QUALITY JOURNALISM IN THE DIGITAL AGE ¶ 2.5.4 (Mar. 17, 2022), https://search.coe.int/cm/pages/result_details.aspx?objectid=0900001680a5ddd0 [https://perma.cc/LT75-LZ3G] (archived Sept. 20, 2022).

156. *Id.*

157. *See* Moriana v. Spain, App. Nos. 36537/15, 36539/15, 2021, (Elósegui, J. & Serghide, J., dissenting), Eur. Ct. H.R. ¶¶ 9–10, https://hudoc.echr.coe.int/eng?i=001-208412 [https://perma.cc/V6SA-9QTQ] (archived Sept. 20, 2022); Rashkin v. Russia, App. No. 69575/10, 2020, (Elósegui, J., dissenting), Eur. Ct. H.R. ¶¶ 6–7, https://hudoc.echr.coe.int/eng?i=001-203811 [https://perma.cc/JY7J-4XPG] (archived Sept. 20, 2022).

158. Commission Recommendation (EU) 2018/334 of 1 March 2018, ch. I, ¶ 4(b), 2018 O.J. (L 63) 8.

159. *See id.* pmbl., ¶¶ 1, 25, 31.

160. *See id.* ch. II, ¶¶ 5–13.

161. *But see* Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), O.J. (L 178), pmbl. ¶¶ 41–48, arts. 14, 15. At the time of writing, the e-Commerce Directive exempts online intermediaries/hosts from liability insofar as they do not have actual

includes not only terrorist content, child sexual abuse material, or illegal hate speech, but also "any information, which, in itself or by its reference to an activity, . . . is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law."[162] For other types of harmful content, including most forms of dis-, mis-, and malinformation, very large platforms would be required to undertake systemic risk assessment and risk mitigation processes, including by giving effect to voluntary codes of conduct,[163] such as the Code of Conduct on online hate speech and the Code of Practice on Disinformation.[164]

Great care is thus needed in calibrating state responses to different IOs. As noted earlier, measures taken by the state are aimed at a speech act, and free speech is itself protected under international human rights law. There are several reasons for caution. First, state regulation of IOs can become a powerful silencing tool in the hands of authoritarian regimes.[165] Second, state regulation that mandates certain rapid assessment and takedown procedures for online intermediaries may relegate decisions impacting human rights to private actors that are ill-suited for this task.[166] Third, overly restrictive sanctions or punishment can have a negative impact on freedom of expression. For platforms, heavy fines and other forms of intermediary liability may drive them to err on the side of content takedowns and other forms of censorship.[167] For individuals, criminal sanctions can have a particularly stigmatizing and chilling effect.[168]

---

knowledge of illegal activity or information or, upon obtaining such knowledge or awareness, act expeditiously to remove or to disable access to the information. The Directive also precludes the imposition of general monitoring obligations on hosts. *See* Communications Decency Act, 47 U.S.C. § 230 (1996). In the United States, intermediary liability for online service providers is provided by US Federal Law.

162. EU Digital Services Act, *supra* note 148, art. 2(g).

163. *See id.* arts. 26, 27, cmts. ¶¶ 63, 68–69.

164. *The EU Code of conduct on countering illegal hate speech online*, COUNCIL OF THE EUR. UNION (Sept. 27, 2019), https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en [https://perma.cc/8XZX-XTS4] (archived Aug. 24, 2022); *Code of Practice on Disinformation*, EUR. COMM'N (Sept. 2018), https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation [https://perma.cc/HT3S-DBUA] (archived Aug. 24, 2022).

165. *See* Caroline Lees, *Fake News – The Global Silencer*, 47 INDEX ON CENSORSHIP 88, 88 (2018).

166. *See* ART. 19, INTERNET INTERMEDIARIES: DILEMMA OF LIABILITY 2 (2013), https://www.article19.org/data/files/Intermediaries_ENGLISH.pdf [https://perma.cc/W2HT-SBS6] (archived Sept. 20, 2022).

167. *See* Monica Horten, *Liability And Responsibility: New Challenges For Internet Intermediaries*, LSE BLOG (Oct. 20, 2016), https://blogs.lse.ac.uk/medialse/2016/10/20/liability-and-responsibility-new-challenges-for-internet-intermediaries/ [https://perma.cc/Z9VS-DZYK] (archived Aug. 24, 2022); Special Rapporteur Report on Disinformation and Freedom of Opinion, *supra* note 20, ¶ 58.

168. *See* Tarlach McGonagle, *"Fake News": False fears or real concerns?*, 35 NETH. Q. HUM. RTS. 203, 204 (2017).

Finally, it is important to note that the application of international human rights law does not depend solely on outlining what positive and negative behaviors it requires of states. Questions of jurisdiction—that is, *where* these obligations apply—also loom large. Jurisdictional issues with respect to human rights have long wrestled with questions of extraterritoriality (i.e., whether and when a state is required to adhere to its positive and negative obligations in areas outside of its territory).[169] Many efforts to apply human rights law to online activities such as IOs have faced similar questions.[170]

When it comes to IOs, however, it is important to emphasize at the outset that the positive and negative obligations reviewed above all clearly apply *within* a state's own territory. For example, ICCPR Article 2(1) provides, "[e]ach State Party to the present Covenant undertakes to respect and to ensure to all individuals within its territory and subject to its jurisdiction the rights recognized in the present Covenant, without distinction of any kind."[171] As a threshold matter, therefore, international law demands respect for human rights by a state vis-à-vis IOs it carries out or effectively controls in its own territory. For those who find themselves in its territory, a state must likewise ensure the protection of those same rights from IOs carried out by other actors based domestically or abroad that do (or will foreseeably) interfere with them.

The situation is more complicated when it comes to IOs affecting individuals abroad. Here, there are a range of views. At one pole lies the United States, which has traditionally suggested its obligations under certain human rights *treaties* have no extraterritorial reach. It reads ICCPR Article 2(1), for example, to only apply to "individuals who are both within the territory of a State Party and subject to that State Party's sovereign authority."[172] The US view, however, is a minority one. And it does not apply to human rights under customary international law, which the United States itself acknowledges apply

---

169. *See* MARKO MILANOVIĆ, EXTRATERRITORIAL APPLICATION OF HUMAN RIGHTS TREATIES: LAW, PRINCIPLES, AND POLICY 222 (2011).

170. *See, e.g.*, TALLINN MANUAL 2.0, *supra* note 14, at 184–86.

171. International Covenant on Civil and Political Rights, *supra* note 76, art. 2(1).

172. *See, e.g.*, U.N. Hmn. Rts. Comm., *Consideration of Reports Submitted by State Parties Under Article 40 o/the Covenant, Third Periodic Reports o/States Parties Due in 2003: United States 0/ America*, annex I, U.N. Doc. CCPR/C/USA/CO/3 (Nov. 28, 2005). In 2010, as Legal Adviser at the US Department of State, Harold Hongju Koh, authored a memorandum arguing that the United States should dispense with this view and apply the obligation to respect human rights extraterritorially; that view, however, did not achieve consensus within the federal government and thus never became US legal policy. *See* U.S. DEP'T OF STATE, OFF. OF THE LEGAL ADVISER, MEMORANDUM OPINION ON THE GEOGRAPHIC SCOPE OF THE INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS 4 (Oct. 19, 2010).

extraterritorially.[173] Recent developments surrounding the conflict in Ukraine further suggest that the restrictive position on the extraterritorial application of international human rights law may be weakening. In April 2022, the UN General Assembly voted to suspend the Russian Federation from the Human Rights Council over "gross and systematic violations and abuses of human rights" committed *in the territory of Ukraine*.[174] The United States voted in favor of this resolution.

Other states and scholars accept an obligation to apply human rights treaty commitments like the ICCPR to areas under a state's effective control even if that area lies outside the state's sovereign territory.[175] A separate view—known as the "personal" model of extraterritorial jurisdiction—has been endorsed by different human rights treaty bodies. It would extend both negative and positive human rights obligations extraterritorially in at least some circumstances where states have physical control or authority over individual rights-holders.[176] Meanwhile, the UN Human Rights Committee and the Inter-American Court of Human Rights have claimed that jurisdiction extends extraterritorially through the activities of entities, such as companies, which are incorporated or located in the state's territory or are otherwise subject to its control. Under this approach, the state's positive duties to protect and ensure human rights extend to the activities of such entities when these have a direct and reasonably foreseeable impact on the human rights of individuals

---

173. *See* NAT'L SEC'Y L. DEP'T, OPERATIONAL LAW HANDBOOK 96 (Maj Ryan Fisher ed. 2022); NAT'L SEC'Y L. DEP'T, OPERATIONAL LAW HANDBOOK 45 (2013) (stating that "[international human rights law] based on [customary international law] binds all States in all circumstances, and is thus obligatory at all times"). For official US personnel (i.e., 'State actors' in the language of IHRL) dealing with civilians outside the territory of the United States, "[customary international law] establishes the human rights considered fundamental, and therefore obligatory." *Id.*; *see also* Ryan Goodman, *The United States' Long (and Proud) Tradition in Support of the Extraterritorial Application of International Human Rights Law*, JUST SEC. (Mar. 10, 2014), https://www.justsecurity.org/8035/united-states-long-and-proud-tradition-supporting-extraterritorial-application-international-human-rights-law/ [https://perma.cc/D7QY-QDJQ] (archived Aug. 24, 2022).

174. *UN General Assembly suspends Russia from the Human Rights Council*, UNA-UK (Apr. 7, 2022), https://una.org.uk/news/un-general-assembly-suspends-russia-human-rights-council [https://perma.cc/MU8C-U2BV] (archived Aug. 24, 2022).

175. *See* Loizidou v. Turkey, App. No. 15318/89, ¶¶ 59–64 (Mar. 23, 1995).

176. *See* Hum. Rts. Comm., General Comment No. 31 [80], The Nature of the General Legal Obligation Imposed on States Parties to the Covenant, U.N. Doc. CCPR/C/21/Rev.1/Add.13, § 10 (May 26, 2004); Coard v. U.S., Case 10.951, Inter-Am. Comm'n H.R., Report N. 109/99, ¶ 37 (1999); MILANOVIĆ, *supra* note 169, at 119 (explaining that the European Court of Human Rights has been reluctant to recognize this model in relation to extraterritorial kinetic force in the absence of governmental control); Marko Milanović, *The Murder of Jamal Khashoggi: Immunities, Inviolability and the Human Right to Life*, 20 HUM. RTS. L. REV. 1, 23–24 (2020).

extraterritorially.[177] Finally, the Human Rights Committee occupies the pole opposite from the United States in interpreting the ICCPR obligations to extend to a state anytime it exercises functional control *over the enjoyment of the rights in question*, regardless of any physical control over territory, the perpetrators, or the individual victim.[178] Hence, pockets of controversy remain around the extraterritorial application of human rights, including when it comes to respecting and protecting certain human rights treaties online. Whether or not a transnational IO falls within the jurisdiction of a state will often depend on the context and the identity of the relevant decision-makers, be they international tribunals and mechanisms, treaty-bodies, non-governmental organizations, or states themselves.

This brief survey suggests four overarching conclusions regarding IOs and international human rights law. First, the protective measures required under this law vary according to the type of content or speech act that the IO features. This requires careful unpacking of the different types of IOs and the various human rights that they implicate. Second, for IOs that involve incitement to violence, for example, states may be under an *obligation* to prohibit them by law. Third, for other types of IOs that may cause harm to certain interests, the state is *entitled to* restrict freedom of speech if the harm affects one of the "legitimate aim" categories provided for under international human rights instruments, but only if the content in question and its restrictive measures are provided by law, necessary, and proportionate in the circumstances. Invasive measures, including content takedowns and sanctions, can only be adopted in accordance with this test. These are without prejudice to the range of other measures that states can—and often must—take to prevent and mitigate the impact of different IOs on various human rights, such as early threat detection, fact-checking, and building awareness and resilience within the population, including through training to detect manipulation. Fourth, as a preliminary matter, States must weigh where the IO occurs. For those occurring within a State's own territory, the application of conventional and customary human rights is straightforward. For

177. *See, e.g.*, HRC General Comment 36, *supra* note 80, § 22; Comm. On Econ., Soc. and Cultural Rts., Statement on the Obligations of States parties regarding the corporate sector and economic social and cultural rights, ¶ 5 U.N. Doc. E/C.12/2011/1 (May 20, 2011); The Environment and Human Rights, Advisory Opinion OC-23/17, Inter-Am. Ct. H.R. (ser. A) ¶¶ 101–02 (Nov. 15, 2017); *see also* Samantha Besson, *Due Diligence and Extraterritorial Human Rights Obligations – Mind the Gap!*, 9 ESIL REFLECTIONS 2, 2 (2020); Milanović & Schmitt, *supra* note 38, at 247, 268.

178. *See* HRC General Comment 36, *supra* note 80, § 63; Georgia v. Russia (II), App. No. 38263/08, ¶¶ 117–44 (Jan. 21, 2021), https://hudoc.echr.coe.int/fre#%7B%22itemid%22:[%22001-207757%22]%7D [https://perma.cc/C4HV-CCNS] (archived Aug. 24, 2022). Despite scholarly support and its recent endorsement by the German Constitutional Court, other human rights bodies have been less enthusiastic about this expansive approach, as evidenced in the Georgia v. Russia (II) judgement of the European Court of Human Rights. *Id.*

those with an extraterritorial nature, competing approaches to jurisdiction under human rights treaties might play out, notwithstanding the universal application of human rights under international custom.

B. *The Principle of Non-Intervention*

In the practice of international relations, states constantly interfere in each other's matters. Despite such frequently observed interferences,[179] it is widely accepted by states that international law contains a binding rule prohibiting coercive intervention in another state's *domaine réservé*.[180] Therefore, the sharp-ended question is not whether this rule exists—it is something that all accept.[181] Rather, the issue is how to understand its contours, particularly as it applies to the ICT context generally and IOs in particular.

Specifying the content of this rule is no easy task. Only eight paragraphs in the *Nicaragua* judgment review non-intervention, and yet this is its most detailed authoritative examination. The principle, in its inter-state form, is not mentioned in the Charter of the United Nations,[182] and, although codified in several regional organizations' constituent instruments (e.g., those of the Organization of American States, the African Union, and the Association of Southeast Asian

---

179. *See* Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), Judgment, 1986 I.C.J. 14, ¶ 186 (June 27). The International Court of Justice in *Nicaragua* did not deem instances of inconsistent behavior to imply an absence of a legal rule: "It is not to be expected that in the practice of States the application of the rules in question should have been perfect, in the sense that States should have refrained, with complete consistency, from the use of force or from intervention in each other's internal affairs. The Court does not consider that, for a rule to be established as customary, the corresponding practice must be in absolutely rigorous conformity with the rule." *Id.* ¶ 126.

180. *See* Mohamed Helal, *On Coercion in International Law*, 52 N.Y.U. J. INT'L L. & POL'Y 1, 65 (2019); HARRIET MOYNIHAN, THE APPLICATION OF INTERNATIONAL LAW TO STATE CYBERATTACKS – SOVEREIGNTY AND NON-INTERVENTION 26 (2018); TALLINN MANUAL 2.0, *supra* note 14, rule 66, ¶ 7; OPPENHEIM'S INTERNATIONAL LAW, VOLUME 1: PEACE § 128 (Robert Jennings & Arthur Watts, eds., 9th ed. 2008).

181. *See* Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security, *Rep. of Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security*, U.N. Doc. A/76/135, ¶ 71(c) (2021) (confirming its existence recently) [hereinafter 2021 GGE Report]; *see also* Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, *Rep. of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, U.N. Doc. A/70/174, ¶ 28 (2015).

182. The Charter does contain a reference to non-intervention in Article 2(7), but this rule is directed at acts of intervention by the United Nations itself. *See* U.N. Charter art. 2(7).

Nations),[183] these treaties do not clarify its elements. Non-intervention featured prominently in the 1970 Friendly Relations Declaration and a number of other resolutions adopted by the UN General Assembly, such as the 1965 Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of their Independence and Sovereignty and the 1982 Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States.[184] However, it is unclear whether each postulate of these non-binding resolutions is reflected in customary international law. Considering the paucity of instruments and judgments that specify and apply the prohibition, Mohamed Helal, for example, has concluded that its contents are "riddled with definitional ambiguity and conceptual uncertainty."[185]

Recent developments, especially states' increasing reliance on ICTs for the conduct of cyber operations, have brought this principle to the forefront of international legal discussions. As such, if adequately specified, this rule can play a meaningful (and constraining) role for harmful state behavior online, including IOs. Attempts at specification are already underway, as many states have proffered their views on the rule in national positions on the application of international law to ICTs.[186]

1. The Starting Point

In *Nicaragua*, the ICJ gave the now-canonical summary of the non-intervention principle, which it found was "part and parcel of customary international law":[187]

> [T]he principle forbids all States or groups of States to intervene directly or indirectly in internal or external affairs of other States. A prohibited intervention must accordingly be one bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely. One of these is the choice of a political, economic, social and cultural system, and the formulation of foreign policy. Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones. The element of coercion, which defines, and indeed forms the very essence of, prohibited

---

183. *See* Charter of the Organization of American States art. 3(e); Constitutive Act of the African Union art. 4(g); ASEAN Charter art. 2(2).

184. *See* G.A. Res. 2625 (XXV) (Oct. 24, 1970); G.A. Res. 2131 (XX) (Dec. 21, 1965); Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States, U.N. Doc. A/RES/36/103 (Jan. 20, 1982).

185. Helal, *supra* note 180, at 47.

186. *See* 2021 GGE Report, *supra* note 181, annex (providing detailed discussions on the principle of non-intervention and reproducing the National positions on the application of international law in cyberspace by Germany (2021), Estonia (2021), Iran (2020), and Norway (2021)); *see also* NEW ZEALAND, DEP'T OF FOREIGN AFFS. AND TRADE, THE APPLICATION OF INTERNATIONAL LAW TO STATE ACTIVITY IN CYBERSPACE ¶ 22 (2020) [hereinafter New Zealand Statement].

187. Nicar. v. U.S., 1986 I.C.J. ¶ 202.

> intervention, is particularly obvious in the case of an intervention which uses force, either in the direct form of military action, or in the indirect form of support for subversive or terrorist armed activities within another State.[188]

In sum, international law prohibits actions attributable to a state (a) involving "methods of coercion" and (b) regarding the internal or external affairs of a state.

2. The Elements

a. Methods of Coercion

What is a method of coercion? While this element was branded by the ICJ as "the very essence" of a prohibited intervention, its indeterminacy is an obstacle to the rule's deterrence pull, as well as its enforcement. There are at least two clusters of questions that are of particular relevance to the IOs observed today.

First, what does the element of coercion tell us about the nature of the rule (i.e., whether it precludes a certain conduct or result)? Neither interpretation would *prima facie* seem unreasonable. The rule could either proscribe a particular wrongful method of engaging in international relations or prohibit the result of having placed another state in a position in which it would not have found itself had it not been for the perpetrator's conduct. In the first variant, conduct dominates; in the second, the outcome is what matters. The former approach is better supported in principle. If the result-based approach is adopted, there is a distinctive risk of excluding an otherwise coercive interference in a state's affairs simply because it proves ineffective.[189] For instance, a deliberate IO to manipulate voters may prove ineffective because of successful measures by the targeted state to increase information literacy within the population. To illustrate, focusing on results might preclude the application of non-intervention to Russian IOs that Ukraine has had success in countering (so far).[190] This is not to say that the outcome of interfering conduct is of no relevance to the inquiry: it could be used as evidence of the ways in which an operation has been conducted.

Second, does the concept of coercion presuppose the existence of a specific subjective attitude of the intervening state, such as a *purpose* to coerce? Relatedly, are there certain methods of coercion that are inherently coercive, irrespective of the intervener's specific intention or purpose? Similarly, short of clear cases of coercion—such as the use or threat of military force and the provision of military support—to

---

188. *Id.* ¶ 205; *see also* Case Concerning Armed Activities in the Territory of the Congo (Dem. Rep. Congo v. Uganda), Judgment, 2005 I.C.J. 168 (Dec. 19).

189. *See* Helal, *supra* note 180, at 43–45.

190. *See* Smith, *supra* note 7.

what extent may other methods amount to coercion? In such cases, including IOs marked by subliminal demands or deception, is the intention to coerce necessary? Can cognitive operations ever amount to coercion?

The easy cases of coercion are dictatorial demands: "Do not do X, or else." In these instances, coercion clearly contains a purposive element. The position of the Netherlands implicitly supports the existence of such purpose, as it defines intervention as "interference in the internal or external affairs of another state *with a view to* employing coercion against that state."[191] Terrorism, insofar as it can be attributed to a state, may be another easy case. Terrorist activity inherently puts forward a forceful demand to its target. It could be argued, however, that a coercive *purpose*—whether specifically proven or inherent in the conduct—is not necessary to satisfy the prohibition of intervention. In the same vein, not just openly coercive methods can compel a state to behave in ways that go against its sovereign will. This wider view, whereby coercion is either intended or foreseeable, could find support in the ICJ's reference to "methods of coercion" in *Nicaragua*. It is also implicit in the ICJ's finding that coercion refers to "choices, which must remain free ones."[192] In short, even without proof of coercive purpose, the very use of methods that can be considered coercive may trigger the prohibition *insofar as the state knew or should have known of their coercive nature*.

Such a broad view of coercion seems to be supported by the United Kingdom, as evidenced in a May 2022 speech by Suella Braverman, then-Attorney General for England and Wales.[193] Specifically, the speech suggests that disruptive cyber behaviors may be "coercive even where it might not be possible to point to a specific course of conduct which a State has been forced into or prevented from taking."[194] For the United Kingdom, "an intervention in the affairs of another State will be unlawful if it is forcible, dictatorial, *or otherwise coercive*, depriving a State of its freedom of control over matters which it is permitted to decide freely by the principle of State sovereignty."[195] Accordingly, the United Kingdom's broader view focuses on certain objective factors as evidence of coercion, such as the scale and effects of an operation.[196] At present, however, it remains to be seen if other states will adopt a similar view.

---

191. Dutch Foreign Ministry Letter to Parliament, *supra* note 107 (emphasis added).

192. Nicar. v. U.S., 1986 I.C.J. ¶ 205.

193. *See* Suella Braverman, International Law in Future Frontiers, Speech at Chatham House (May 19, 2022), https://www.ukpol.co.uk/suella-braverman-2022-speech-at-chatham-house/ [https://perma.cc/7ZSJ-9XST] (archived Aug. 24, 2022).

194. *Id.*

195. *Id.* (emphasis added).

196. *See id.*

In any case, there are sound reasons to adopt a broader, and more objective approach to the element of coercion. First, the severity of certain types of harms risked or caused by cyber and information operations (e.g., disruption of emergency medical services, electoral processes, or power generation infrastructure) justifies their designation as unlawful intervention regardless of any underlying purpose. Second, a purpose in deploying an IO may be very difficult, if not impossible, to ascertain in the case of abstract entities such as states. Third, a narrow definition of coercion might incentivize states to pursue *more* online operations that, even if not launched with an intention to coerce, could be characterized by recklessness or negligence vis-à-vis their harmful effects.

Thus, the better interpretation of "coercion" is one that is (1) conduct-based and (2) focused on the employment of wrongful methods that may foreseeably place a state in a coerced position, compelling it to behave in unwanted ways. This, in turn, leads to additional questions. Which methods are wrongful? Is an IO campaign in an election not coercive because, by definition, it seeks only to persuade a population (or certain decision-makers)? Or can it constitute coercion because of its ultimate aim to unsettle the targeted political system? Some scholars seem to accept the latter line of thinking, arguing that the provision of false information to voters—as opposed to state agents themselves—should be considered coercive.[197] Steven Wheatley, for instance, argues that coercion "describes a situation in which the outside power forces the target to do something they would not otherwise do," and that one way of creating this situation is to lie "to the electorate with the intention of deceiving them into thinking and then voting differently."[198] A standard of untruth is unpersuasive as an appropriate benchmark, however, since even truthful pieces of information, arranged in particular ways and delivered at an appropriate time, can be highly manipulative. That said, the underlying principle is sound: certain types of IOs may qualify as methods of coercion insofar as they intentionally or foreseeably deprive the target state of its sovereign will. This is particularly the case where deception is present.

---

197. *See* Steven Wheatley, *Cyber and Influence Operations Targeting Elections: Back to the Principle of Non-Intervention*, EJIL: TALK! (Oct. 26, 2020), https://www.ejiltalk.org/cyber-and-influence-operations-targeting-elections-back-to-the-principle-of-non-intervention/ [https://perma.cc/3AMP-NKJU] (archived Aug. 24, 2022); Nicholas Tsagourias, *Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace*, EJIL: TALK! (Aug. 26, 2019), https://www.ejiltalk.org/electoral-cyber-interference-self-determination-and-the-principle-of-non-intervention-in-cyberspace/ [https://perma.cc/3S6Y-JUT4] (archived Aug. 24, 2022).

198. Wheatley, *supra* note 197.

In short, the element of coercion is a treasure chest of questions with, as of yet, few conclusive answers. States are skating on thin ice when navigating this element. Interpreted too broadly, the element risks collapsing on itself. Interpreted too narrowly, it may exclude certain types of conduct that are deemed unacceptable in international relations. Ultimately, coercion's role should be to separate acceptable acts of persuasion, influence, criticism, and public diplomacy from wrongful interference, as understood in the age of endemic campaigns of disinformation, deception, and disruption.

b. Internal or External Affairs

Intervention only exists when state conduct constitutes coercion regarding another state's "internal or external affairs."[199] In other words, coercive acts will not trigger the prohibition of non-intervention absent a connection to the affairs of a state. Thus, even if state-sponsored ransomware is, by definition, coercive (e.g., demanding payment in return for restoring access to a system, network, or data), it will only be prohibited where it concerns the state's internal or external affairs.[200]

Equating internal and external affairs to a state's *domaine réservé* has done little to resolve the definitional difficulties.[201] States have yet to decide if the *domaine réservé* is dynamic or immutable. Is the *domaine réservé* defined by some sort of fixed (and objective) list of inherently sovereign functions? Or does it fluctuate on the basis of international obligations undertaken by a state? If the latter is true, then the *domaine réservé* is a residual category, which can be eroded by the will of the state whose affairs it concerns. This does not mean, however, that undertaking a certain international obligation in one area, such as health, necessarily excludes it from a state's *domaine réservé*. For, even in those instances, states retain a wide margin of discretion in the implementation of their international obligations.[202]

That said, there is room for states to identify non-intervention's protected subjects beyond those listed in existing doctrine (e.g., the "choice of a political, economic, social, and cultural systems, and the formulation of foreign policy").[203] Discussions among the UN Group of Governmental Experts on Advancing Responsible State Behaviour in

199. Nicar. v. U.S., 1986 I.C.J. ¶ 205.

200. *See id.*

201. *See* TALLIN MANUAL 2.0, *supra* note 14, at 314. *But see* MOYNIHAN, *supra* note 180, at 33–34 (arguing that the duty of non-intervention protects a state's "inherently sovereign functions" rather than the *domaine réservé* that involves a sphere of activity that is not otherwise regulated by international law).

202. *See* Priya Urs, The Application of the Prohibition of Intervention to Cyber Operations against the Healthcare Sector 12 (May 2022) (Draft Research Paper, Oxford Institute for Ethics, Law and Armed Conflict) (copy on file with authors).

203. Nicar. v. U.S., 1986 I.C.J. ¶ 205.

Cyberspace in the Context of International Security revealed some common candidates[204]—including the information ecosystem involving electoral processes and campaigns, critical infrastructure, and medical facilities—that have already been the target of IOs.[205] This is of direct relevance to the regulation of IOs, as these may target a state's conduct of elections or its pandemic response.

c. The Evolution of the Non-Intervention Principle

Non-intervention is widely recognized as a corollary to every state's sovereignty. But what does this mean? Is it related to a *principle* of sovereignty with no substantive content of its own, or is the existence of non-intervention tightly connected to a self-standing rule of sovereignty? And if the latter is true, how do the two rules interact? At present, there are no clear answers. As discussed below, some states accept the existence of a standalone rule of sovereignty that may also regulate IOs that states pursue. Such states may interpret the element of coercion more narrowly, while those who object to it may adopt a broader reading.

The evolution of non-intervention may also be closely linked with the prohibition of the threat and use of force. In the *Nicaragua* judgment, the two rules were tied together through the element of coercion: as the excerpt above provides, coercion "is particularly obvious in the case of an intervention which uses force."[206] Threats of force may also be paradigmatic examples of coercive behavior in a state's internal or external affairs, as they seek to compel a change in their addressee's domestic sphere. Because of these close ties, certain

---

204. Pursuant to the mandate contained in General Assembly Resolution 73/266, the 2019–2021 Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security was tasked with the study, with a view to promoting common understandings and effective implementation, of possible cooperative measures to address existing and potential threats in the sphere of information security. Its July 2021 Report contained the Group's findings on "existing and emerging threats; norms, rules and principles for the responsible behaviour of States; international law; confidence-building measures; and international cooperation and assistance in ICT security and capacity-building. On each of these topics, the report adds a layer of understanding to the findings and recommendations of previous Groups of Governmental Experts." 2021 GGE Report, *supra* note 181, at Summary.

205. On elections, see generally 2021 GGE Report, *supra* note 181, annex. Australia, Brazil, Estonia, Germany, Netherlands, Norway, Romania, Singapore, United Kingdom, and United States all list elections as among the affairs protected by the non-intervention principle with several—e.g., Germany, Norway—including disinformation campaigns. States like Estonia, Norway, Japan, and New Zealand have expressed support for protecting critical infrastructure, while others like Japan, New Zealand, the United Kingdom, and the United States would also add medical services (with New Zealand including disinformation campaigns that significantly undermine a state's public health efforts during a pandemic). *See id.;* New Zealand Statement, *supra* note 186, ¶ 10.

206. Nicar. v. U.S., 1986 I.C.J. ¶ 205.

elements of the prohibition of threat or use of force may carry over into the non-intervention discussions. For instance, the purported threshold of "scale and effects" in the use of force is already lurking in Germany's position on non-intervention, which posits that, for a cyber measure to constitute a prohibited intervention, it must be comparable in scale and effect to coercion in non-cyber contexts.[207] Applying this approach to foreign electoral interference, Germany opines that

> a State, by spreading disinformation via the internet, may deliberately incite violent political upheaval, riots and/or civil strife in a foreign country, thereby significantly impeding the orderly conduct of an election and the casting of ballots. Such activities may be comparable *in scale and effect* to the support of insurgents and may hence be akin to coercion in the above-mentioned sense.[208]

There is thus a clear role for the principle of non-intervention in the regulation of IOs. Progress in the clarification of the rule may be slow but is undeniable. The regulation of IOs may become the perfect testing ground for applying non-intervention in an ICT context. The complex and sophisticated IOs we see today leave no space for avoiding the difficult questions and states' recent engagement with the rule shows a renewed willingness to engage on these issues.

C. *Sovereignty*

Without any doubt, sovereignty is a foundational and organizing principle in international law. Sovereign equality lies at the heart of the international legal system.[209] Sovereignty guards territorial boundaries and ensures the exclusive legislative, administrative, and judicial competence of states over their airspace.[210] It is also closely related to the principle of non-intervention, with the latter seen as a manifestation of state sovereignty over the *domaine réservé*.[211] In recent years, a number of states have asserted in the context of ICTs that respect for sovereignty is a self-standing rule of international law—the breach of which, when attributable to a state, would constitute an internationally wrongful act.[212] Most recently, Canada's

---

207. *See* FED. GOV'T OF GER., ON THE APPLICATION OF INTERNATIONAL LAW IN CYBERSPACE: POSITION PAPER 5 (Mar. 2021), https://www.auswaertiges-amt.de/blob/2446304/32e7b2498e10b74fb17204c54665bdf0/on-the-application-of-international-law-in-cyberspace-data.pdf [https://perma.cc/K5SF-AZ88] (archived Aug. 24, 2022).

208. *Id.* (emphasis added).

209. *See* U.N. Charter art. 2(1).

210. *See id.* art. 2(4); Convention on International Civil Aviation art. 1, Dec. 7, 1944, 15 U.N.T.S. 295.

211. The recent speech by the UK Attorney General sees sovereignty and non-intervention as two sides of the same coin. *See* Braverman, *supra* note 193.

212. *See, e.g.*, Dutch Foreign Ministry Letter to Parliament, *supra* note 107; *see also* MINISTRY OF DEFENSE OF FRANCE, INTERNATIONAL LAW APPLIED TO OPERATIONS IN

April 2022 national position on the application of international law to cyberspace dedicates twelve paragraphs to the contours of sovereignty as a rule.[213] A month and a half later, the United Kingdom reaffirmed its view that there is insufficient evidence to extrapolate a rule of sovereignty from the principle.[214] As such, the existential debate over sovereignty remains very much in flux.

For proponents of a self-standing rule of sovereignty, it provides two distinct avenues for protection against harmful cyber operations. According to Michael N. Schmitt and Marko Milanović, "the sovereignty of a State may be breached by cyber operations attributable to another State in two basic ways—by causing effects on the territory of the former or by interfering with its inherently governmental functions, even in the absence of territorial effects."[215] On the first way, the *Tallinn Manual 2.0* posits that relatively permanent interference with the functionality of cyber infrastructure would qualify as "effects" or "consequences" for the purposes of the rule.[216] On the second, interference with, or usurpation of, an inherently governmental act triggers the rule. The conduct of elections is a paradigmatic example of such a function. And while other examples like healthcare or cybersecurity are not necessarily governmental functions across jurisdictions, crisis management and national security, including in the context of infectious diseases, would qualify.[217]

Part of the appeal of the rule of sovereignty is that it may circumvent the difficulties associated with defining "coercion" in the rule of non-intervention. However, sovereignty-as-a-rule comes with its own challenges. Two are worth flagging in particular. First, as seen earlier, sovereignty's very existence as a self-standing rule of international law is still in doubt.[218] Second, the boundaries of the rule are equally unclear. In 2020, the general counsel of the US Department

---

CYBERSPACE 6–7 (Sept. 9, 2019) [hereinafter French Position on International Law Applications to Cyberspace]; Video Tape: Statements of Austria, Finland and the Czech Republic at the 2nd Substantive Session of Open-Ended Working Group Working Group on developments in the field of information and telecommunications in the context of international security (2020).

213. *See International Law applicable in cyberspace*, GOV'T OF CAN., https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/peace_security-paix_securite/cyberspace_law-cyberespace_droit.aspx?lang=eng [https://perma.cc/968R-A95G] (archived Aug. 24, 2022).

214. *See* Braverman, *supra* note 193.

215. Milanović & Schmitt, *supra* note 38, at 253.

216. TALLINN MANUAL 2.0, *supra* note 14, at 20–21.

217. *See* Milanović & Schmitt, *supra* note 38, at 255.

218. *See, e.g.*, Jeremy Wright, Att'y Gen. of the U.K., Address at Chatham House, Cyber and International Law in the 21st Century (May 23, 2018) (noting that the United Kingdom is opposed to sovereignty-as-a-rule and has forcefully rejected its existence for a number of years) (transcript available at U.K. Attorney General's Office).

of Defense, Paul Ney, opined that not "all infringements on sovereignty in cyberspace necessarily involve violations of international law."[219] In contrast, France takes a very broad view of this rule:

> *[A]ny cyberattack against French digital systems* or *any effects produced* on French territory by digital means by a State organ, a person or an entity exercising elements of governmental authority or by a person or persons acting on the instructions of or under the direction or control of a State constitutes a breach of sovereignty.[220]

Assuming the existence of this rule within the parameters outlined by Schmitt and Milanović, it has the potential to play a key role in restraining the conduct of inter-state IOs. For instance, IOs that mislead the public of another state into ingesting a dangerous "miracle cure" during a pandemic, risking illness and death, could qualify as breaches of sovereignty under the "territorial effects" limb. It is equally easy to imagine IOs that have the capacity to interfere with inherently governmental functions, such as operations misinforming voters on ways to cast their votes in electoral processes or spreading false information on the availability of vaccination sites in a given area.

Whether through the establishment of a self-standing rule of sovereignty or an expanded interpretation of the principle of non-intervention, it seems clear that states are showing an increased willingness to consider a broad array of harmful online activities as covered under the *lex lata* corpus of international law. While IOs as a category may not always explicitly feature in the national positions of states, the elements of the rules outlined under the banners of non-intervention and sovereignty clearly capture a range of such operations on the basis of their effects in the territory or institutions of another state.

D. *Due Diligence Standards: The* Corfu Channel *and the No-Harm Principles*

Different types of IOs may also fall within the scope of two related but distinct rules requiring states to exercise due diligence in preventing, halting, and/or redressing certain harms.

The first of these rules is the *Corfu Channel* principle, which borrows its name from the very first case decided by the ICJ in 1949 between the United Kingdom and Albania. There, the ICJ found that it is a "well-recognized principle" that every state has an "obligation not to allow knowingly its territory to be used for acts contrary to the

---

219. Paul C. Ney, Jr., Dep't of Def. Gen. Couns., DOD General Counsel Remarks at U.S. Cyber Command Legal Conference (Mar. 2, 2020) (transcript available at U.S. Department of Defense).

220. French Position on International Law Applications to Cyberspace, *supra* note 212, at 3 (emphasis added).

rights of other States."[221] To the extent that internal or cross-boundary IOs may be carried out by states or non-state actors and contravene the victim state's sovereign rights, including the rights to non-intervention, self-determination, or the human rights of its population, they may well be covered by the *Corfu Channel* principle. This means that a state from whose territory or physical infrastructure the IO is carried out (or through which it transits) must exercise its best efforts to prevent or stop the operation from undermining the rights of other states, whether or not the operation can be attributed to the duty-bearer.[222]

This duty does not require states to successfully prevent or stop all such operations or acquire knowledge thereof. Rather, it requires states to put in place the minimal governmental infrastructure and feasible measures with a view to preventing or halting operations of which it knows or should have known.[223] Although the *Corfu Channel* principle has a chiefly preventative outlook, the obligation is only breached if the actual harm—an act contrary to the rights of other states—materializes.[224] Moreover, as an obligation of general applicability to states under international law, the *Corfu Channel* principle applies across different domains and technologies, including "cyberspace" or ICTs.[225] Examples of IOs potentially falling within the scope of the *Corfu Channel* principle include disinformation campaigns that undermine a state's right to exercise sovereign functions, such as the conduct of elections and health crisis management.

The second rule featuring a due diligence standard and binding states under international law is the no-harm principle. This obligation requires states to prevent, stop, and redress significant transboundary "harm to persons, property *or* the environment," irrespective of attribution of this conduct to a state.[226] Unlike the *Corfu Channel* principle, the no-harm principle also covers activities *not* prohibited under international law.[227] Its scope encompasses cross-boundary events causing significant harm (i.e., something more than detectable), leading to "a real detrimental effect on matters such as, for example, human health, industry, property, environment or agriculture in other States."[228] Such effects must be measured by factual and objective standards.[229] What qualifies as significant will depend on value

---

221. Corfu Channel (U.K. v. Alb.), Judgment, 1949 I.C.J. 4, at 22 (Apr. 9).
222. *See* Coco & de Souza Dias, *supra* note 53, at 783–84.
223. *See id.* at 789.
224. *See* ARSIWA, *supra* note 56, art. 14(3).
225. *See* Coco & de Souza Dias, *supra* note 53, at 778–83.
226. Int'l L. Comm'n, Rep. on the Work of Its Fifty-Third Session, U.N. Doc. A/56/10, at 152–53, art. 2(b), cmt. ¶¶ 8–9 [hereinafter ILC Draft Articles on Prevention].
227. *See id.* at 150, cmt. to art. 1, ¶ 6.
228. *Id.* at 152, cmt. to art. 2, ¶ 4.
229. *See id.*

determinations and the circumstances prevailing at the time, such as the available scientific evidence, human appreciation for a certain object, and the probability and magnitude of the harm.[230]

Failure to exercise due diligence under the no-harm principle gives rise to *liability* to compensate the harm once it materializes.[231] It is only when this liability is not met through compensation or other forms of redress that international responsibility arises.[232] There is no question that the no-harm principle applies beyond the ecological context, with the International Law Commission (ILC)'s work on transboundary harm extending it to "all physical uses of territory giving rise to adverse physical transboundary effects."[233] Yet controversy remains as to whether the no-harm principle is limited to physical harm or extends to cover non-physical harm, such as moral, financial, and reputational damage.[234]

We submit that the principle does apply to non-physical harms, such as those caused by certain IOs, for three main reasons. First, the decision to limit the scope of the ILC's work to the *physical consequences* of transboundary harms was a purely pragmatic one.[235] Second, as in the context of human health, it has become increasingly difficult to distinguish between physical and non-physical harms in the digital age. This is partly due to the interconnectedness between physical and digital systems, such as hardware and software, as well as the importance of data and its processing for daily life. Third, evidence surveyed by the ILC itself suggests that the customary formulation of the no-harm principle applies to *all* kinds of transboundary harm to persons, property, or the environment caused by activities carried out within a state's jurisdiction or territory. There are various examples of state practice and/or *opinio juris* in support of

---

230. *See id.* at 152–53, cmt. to art. 2, ¶¶ 3, 7.

231. *See id.* at 148 (General Cmt.), ¶ 1 at 150, cmt. to art. 1, ¶ 6.

232. *See* Coco & de Souza Dias, *supra* note 53, at 794.

233. Robert Q. Quentin-Baxter (Special Rapporteur on International Liability for Injurious Consequences Arising out of Acts Not Prohibited by International Law), *Fourth report on international liability for injurious consequences arising out of acts not prohibited by international law*, ¶ 17, U.N. Doc. A/CN.4/373 + Corr.1–2 (June 27, 1983).

234. *See* ILC Draft Articles on Prevention, *supra* note 226, at 151, cmt. to art. 1, ¶ 16.

235. *See id.*; Pemmaraju Sreenivasa Rao (Special Rapporteur on International Liability for Injurious Consequences Arising out of Acts Not Prohibited by International Law), *Liability regimes relevant to the topic "International liability for injurious consequences arising out of acts not prohibited by international law"*, ¶¶ 253–271, U.N. Doc. A/CN.4/471 (June 23, 1995) [hereinafter 1995 Survey of Liability Regimes]; *see also* U.N. Secretariat, Survey of liability regimes relevant to the topic of international liability for injurious consequences arising out of acts not prohibited by international law (international liability in case of loss from transboundary harm arising out of hazardous activities), ¶¶ 526–30, U.N. Doc. A/CN.4/543 (June 24, 2004).

this assertion.[236] These include various treaty provisions applicable in the ICT context, such as (a) Article 10(2) of the 1927 International Radiotelegraph Convention,[237] (b) Article 35(1) of the 1932 International Telecommunication Convention,[238] and (c) Articles 38(5) and 45(3) of the 1992 Constitution of the International Communications Union.[239]

Articles 1 to 4 of the 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace are of particular importance for the regulation of IOs. The convention, specifically tailored for the then-prevalent phenomena of propaganda and false news, requires states parties to, *inter alia*, prohibit and stop the broadcasting from their territory of any transmission that (i) "is of such a character as to incite the population of any territory to acts incompatible with the internal order or the security of a territory of another party" *and* (ii) is "likely to harm good international understanding by Statements the incorrectness of which is or ought to be known to the persons responsible for the broadcast."[240] It also requires states parties "to ensure, especially in time of crisis, that stations within their respective territories shall broadcast information concerning international relations the accuracy of which shall have been verified . . . by the persons responsible for broadcasting the information."[241]

In the context of IOs, non-physical harms possibly covered by the no-harm principle include reputational damage arising from dis-, mis, and malinformation campaigns, whether carried out by states or non-state actors. Likewise, online hate speech could cause direct psychological or moral harm to its individual addressees, instigate violence against particular persons, and lead to broader political instability and unrest in states. Controversies aside, it is beyond doubt that states must prevent, halt, and redress significant transboundary harm or risk to the life, health, or physical integrity of individuals

---

236. *See* U.N. Secretariat, Survey of State practice relevant to international liability for injurious consequences arising out of acts not prohibited by international law, ¶¶ 55–69, U.N. Doc. A/CN.4/384, (Oct. 16, 1985). The ILC considered those agreements it listed to be a reflection of "the general requirement that States must assess the injurious impact of activities undertaken by them or by persons under their control," as stated in judicial decisions such as Trail Smelter and Corfu Channel. *Id.* ¶ 65.

237. *See* International Radiotelegraph Convention art. 10(2), Nov. 25, 1927, 45 Stat. 2843, T.S. No. 767.

238. *See* 1932 International Telecommunication Convention art. 35(1), Dec. 9, 1932, 49 Stat. 2393, T.S. No. 867.

239. Constitution and Convention of the International Communications Union arts. 38(5), 45(3), Dec. 22, 1992, 1825 U.N.T.S. 33.

240. International Convention concerning the Use of Broadcasting in the Cause of Peace arts. 1, 3, Sept. 23, 1936, 186 U.N.T.S. 301 [hereinafter 1936 Broadcasting Convention].

241. *Id.* art. 4.

caused by IOs. As such, IOs constituting incitement to discrimination, hostility, or violence are clearly covered by these rules.

IV. DOES INTERNATIONAL LAW NEED CLARIFICATION OR DEVELOPMENT WITH RESPECT TO IOS?

The foregoing survey reveals a robust array of international obligations that extend to state behavior vis-à-vis IOs. International human rights law, non-intervention, sovereignty, and due diligence obligations all impact what IOs a state can conduct and what actions states must pursue when others do so. Yet saying international law governs IOs does not tell us much about how well it does so. Turning from the law's regulation to its efficacy, there are no less than four areas where the extant law comes under some strain—in its (i) application, (ii) orientation, (iii) complexity, and (iv) enforcement. Such challenges may help explain why, despite all the existing law, we are witnessing a rise in harmful IOs by a range of international actors from states to non-state terrorist networks.

A. *Application Problems*

International law's operation depends, first and foremost, on having the relevant facts for its application. When it comes to IOs, however, the relevant facts are often shrouded by issues of identification, assessment, and attribution. For starters, the default nature of IOs is usually secrecy or obfuscation. For most IOs, the planning, design and execution are not meant to become public.[242] Yet, to impose legal responsibility for state behavior, it is necessary to identify the relevant behavior in the first place. As such, many IOs may proceed without their targets ever knowing their nature or origins, let alone in a sufficiently timely manner so as to allow for some response. Indeed, there are few metrics on baseline conditions today (i.e., an understanding of whether most IOs conducted by states become publicly known remains in the shadows). And even where there is some public behavior, such as claims that the 2020 US presidential election was "stolen," the fact that such behavior was an orchestrated IO will always be denied.[243] If the nature and origin of an IO are unknown, it

242. *See* Herbert Lin, *Conclusion: An Outsider Looks In, in* DEFENDING DEMOCRACIES, *supra* note 18, at 368 (highlighting that, in the context of IOs targeting foreign elections, "proxy accounts, domestically based 'useful idiots', and accounts established long before the election" make recognizing such IOs "difficult and time-consuming").

243. *See* DIEGO A. MARTIN, JACOB N. SHAPIRO, & JULIA G. ILLHARDT, EMPIRICAL STUD. OF CONFLICT PROJECT, ONLINE POLITICAL INFLUENCE EFFORTS DATASET 9 (2020), https://esoc.princeton.edu/publications/trends-online-influence-efforts [https://perma.cc/GHE9-DHLM] (archived Aug. 24, 2022) (tracking 114 distinct influence efforts

becomes difficult to apply law to it at all, let alone to do so effectively. Similar issues arise with respect to the transparency of private sector (e.g., social media) responses to the use of their platforms for IOs, which many companies have chosen to avoid disclosing, perhaps to improve their efficacy or sustain their advertising business models.

And even where victims (or others) identify an IO, they are likely to encounter particular problems assessing it. Simply put, it is not enough to know that some IO is ongoing (or has already happened). International lawyers also need to have a sense of what the IO does (or is capable of doing). In international humanitarian law, for example, international lawyers will assess an attack in terms of the means used (i.e., whether they are capable of distinguishing between civilians and civilian objects, on the one hand, and combatants and military objects, on the other) and its expected effects (i.e., whether the expected incidental civilian harm was proportionate to the military advantage anticipated) just as human rights lawyers must identify when state conduct constitutes torture or incitement to violence. On occasion, similar assessments may be possible with respect to certain IOs. It is not difficult to assess public laws enacted by states to fulfill (or remain within the confines) of their human rights obligations. And although slightly harder, we can envision a similar capacity to assess whether a ransomware operation was prohibited by international law based on its targeting of information infrastructure associated with inherently governmental functions (e.g., election websites or public utilities).

Assessing the means, scale, and effects of most IOs, however, presents a much higher hurdle. Accurately mapping what an IO does (i.e., how it navigates to and through its targets and achieves its end goal) can be quite difficult empirically. Moreover, when individual elements of an IO are considered in isolation, they may appear lawful even as they cumulatively compile into an internationally wrongful outcome, a process that may be difficult to measure even with the benefit of hindsight.[244] On top of this, significant questions of causation remain; at present, international lawyers often lack the research tools to know when IOs work (i.e., when are they just "noise" and when do they actually change minds (or votes) or otherwise increase the likelihood that their targets will act in ways the IO's authors intend).[245] Causation issues are not, of course, unique to IOs—the import of other

---

from 2011 to 2021 that targeted fifty-six countries; eighty-four influence efforts were attributed in origin to foreign states, while thirty influence efforts can be attributed to a government influencing its own domestic population).

244. *See* Lin, *supra* note 242, at 374 ("Law does not deal well with activities that are tolerable in small numbers but intolerable in large numbers, a problem that is exacerbated by the rise in information technologies . . . the amplification of messages that are likely quite harmless in onesies and twosies but potentially quite harmful when distributed on scales larger by several orders of magnitude.").

245. *See id.*

state behaviors (e.g., economic sanctions) is often unclear. Yet we believe these issues are more acute in the IO context given the cognitive aspirations of the behavior in question and the fact that it is often the case that, for an IO to be unlawful, someone will often still need to act upon it.

Beyond identification and assessment, issues of attribution loom large. As noted, attribution necessarily involves determining the IO's authors (or at least its point of origin). Even if activity could be linked to a particular IP address or computer network, the use of botnets, VPNs, proxies, and other tactics, techniques, and procedures (TTPs) mean that the operator behind the IO could remain anonymous, to say nothing of the identity of those who directed or controlled that individual.[246] These problems have receded in recent years, at least for a handful of the most highly skilled and resourced states (as well as some of the Big Tech companies themselves), which now have greater visibility into the technical origins of cyber-operations.[247] In some cases, the use of learning algorithms for pattern recognition may allow associations of particular TTPs to specific threat actors, meaning that it is possible to assign political responsibility to a state or organization even if there is insufficient direct evidence to tie the underlying IO to the particular individuals who perpetrated them.[248] Such advances, however, are not universal; they are uneven, as many states (to say nothing of other actors like international organizations) lack the capacity to perform technical attribution at all. And even those who have such skills may not always meet with success consistently as IO authors continue to innovate the ways of hiding their operations and identities.

These factual-application issues are exacerbated when paired with equally thorny issues of *how* the laws themselves apply. As already noted, in several areas, states face "existential" questions about whether a specific international law obligation applies in the ICT environment at all.[249] Issues as to whether sovereignty exists as a standalone rule in cyberspace or if due diligence comprises one or more legally binding duties—or none at all—have obvious implications for

---

246. *See* Jon R. Lindsay, *Tipping The Scales: The Attribution Problem And The Feasibility Of Deterrence Against Cyberattack*, 1 J. CYBERSECURITY 53, 54 (2015).

247. *See, e.g.*, Benjamin Edwards, Alexander Furnas, Stephanie Forrest & Robert Axelrod, *Strategic Aspects Of Cyberattack, Attribution, And Blame*, 114 PROC. NAT'L ACAD. SCIS. 2825, 2825 (2017); JOHN S. DAVIS II, BENJAMIN BOUDREAUX, JONATHAN WILLIAM WELBURN, JAIR AGUIRRE, CORDAYE OGLETREE, GEOFFREY MCGOVERN & MICHAEL S. CHASE, STATELESS ATTRIBUTION: TOWARDS INTERNATIONAL ACCOUNTABILITY FOR CYBERSPACE 1, 2 (Rand Corp. 2017).

248. *See* HEALEY, *supra* note 52, at 2.

249. *See* Duncan B. Hollis, *The Existential Function of Interpretation in International Law*, *in* INTERPRETATION IN INTERNATIONAL LAW 90 (A. Bianchi et al. eds., Oxford Univ. Press, 2015) (on the difference between existential and other forms of interpretative disputes).

the law's application to IOs. And even where states and other stakeholders concede a rule's existence (e.g., international human rights and the duty of non-intervention), states have yet to delineate what these rules mean in relation to the use and effects of behaviors online. The extraterritorial application question hampers the application of certain human rights treaties abroad. Inter-state obligations, such as non-intervention, may prohibit certain IOs, just as the *Corfu Channel* and no-harm principles may require states to exercise due diligence to prevent, end, or redress them. But interpretative challenges over which IOs so qualify remain, leaving uncertainty and ambiguity over the law's application in their wake.

On top of all this, states appear to still be divided on what process(es) to use to redress these existential and interpretative issues. Over the last few years, calls for more transparency on how states understand international law as it applies in the ICT context have generated a rising number of national statements that offer official views.[250] However, these statements have yet to generate clear demarcations of areas of convergence (and divergence) on the law's application for those making them. Meanwhile, a majority of states have yet to express any views. In practice, moreover, states rarely refer to international law explicitly even as they increasingly attribute cyber operations, including IOs, to foreign state actors in individual and collective accusations.[251]

B. *Orientation Problems*

International law was not designed for a digital world. As such, it fits unevenly with the rising number of IOs and related cyber activities. There are few tailor-made international rules for IOs, none of which were designed to accommodate the internet's capacities and reach.

---

250. *See, e.g.*, 2021 GGE Report, *supra* note 181, annex (collecting national views from Australia, Brazil, China, Estonia, France, Germany, India, Indonesia, Japan, Jordan, Kazakhstan, Kenya, Mauritius, Mexico, Morocco, Netherlands, Norway, Romania, Russian Federation, Singapore, South Africa, Switzerland, United Kingdom, United States, and Uruguay); Duncan B. Hollis, *Fifth Report - International Law and State Cyber Operations: Improving Transparency*, Inter-American Juridical Committee, 97th Regular Session, Organization of American States, CJI/doc. 603/20 (July 17, 2020) (surveying official, national responses to a questionnaire from Bolivia, Brazil, Chile, Costa Rica, Ecuador, Guatemala, Guyana, Peru, and the United States) [hereinafter Hollis, *Fifth Report*].

251. *See, e.g.*, Finnemore & Hollis, *supra* note 50, at 971–72; Dan Efrony and Yuval Shany, *A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyber-operations and Subsequent State Practice*, 112 AM. J. INT'L L. 583, 594 (2018). In one notable exception, several states did invoke international law as part of the coordinated accusation that the GRU—Russia's military intelligence arm—was responsible for a series of cyber-operations, including those targeting the Organization for the Prohibition of Chemical Weapons (OPCW) and the World Anti-Doping Agency (WADA). Finnemore & Hollis, *supra* note 50, at 990.

Those rules that do exist are specific to particular contexts (e.g., incitement to discrimination, hostility, or violence,[252] or false statements that "harm good international understanding"[253]). Even in those specific contexts, existing regulations are often ad hoc and piecemeal. The specific international legal framework applicable to terrorism provides a prime example. International laws that protect against terrorism have emerged across over a dozen treaties that regulate specific ends and means that may implicate an IO. An IO, for example, may incentivize behavior that leads to a terrorist bombing.[254] But the law chiefly regulates the bombing itself rather than the IOs' catalyzing steps. Other existing rules emphasize kinetic (e.g., using plastic explosives[255]) or economic (financing[256]) methods that also operate at some distance from an IO itself. As such, even while these rules may occasionally overlap with an IO and regulate some of its aspects or effects, gaps abound.

More fundamentally, international law offers limited focus on IO's "cognitive" orientation. Most international law regulates "physical" behavior (i.e., people, places, and things). The internet brought with it a virtual dimension of "information" communications at a scale never seen before (i.e., the massive collection, processing, storage, and dissemination of data by digital systems).[257] When it comes to IOs, however, these operations implicate a different, "cognitive" set of behaviors, focused primarily on human minds, attitudes, and emotions. These three dimensions of behavior—physical, informational, and cognitive—are inter-related. Physical behaviors and information communications can influence what people think and feel, just as what people think and feel may lead to specific communications or physical behavior. Yet asking international law to regulate physical and information activities with an eye to their cognitive causes or consequences is a different, and undoubtedly far more difficult, task than asking it simply to regulate those behaviors alone.[258] This challenge is compounded by the paucity of (and confusion over) general standards of causation for internationally wrongful

---

252. *See* International Covenant on Civil and Political Rights, *supra* note 76, art. 20(2).

253. 1936 Broadcasting Convention, *supra* note 240, arts. 1, 3.

254. *See* International Convention for the Suppression of Terrorist Bombings, Dec. 15, 1997, 2149 U.N.T.S. 256.

255. Convention on the Marking of Plastic Explosives for the Purpose of Detection, March 1, 1991, 2122 U.N.T.S. 359.

256. International Convention for the Suppression of the Financing of Terrorism, Dec. 9, 1999, 2178 U.N.T.S. 197.

257. This addition has generated its own set of questions regarding the law's application and efficacy, including issues of data protection that spawned the EU General Data Protection Regulation and various other regional and national approaches. *See, e.g.*, Parliament Regulation 2016/679, 2016 O.J. (L 119).

258. *See* Lin & Kerr, *supra* note 36, at 269.

acts.[259] Regulating to affect what people may (or may not) say or think may not even work. It is also unclear whether the international community even wants to regulate cognitive methods, since (as discussed below) it may open the door to harmful implications of its own.

But the challenge for international law does not end simply by covering IOs via pre-existing general (and occasionally specific) obligations that regulate physical or informational acts. International law has a statist orientation—it provides a set of prohibitions, permissions, and requirements *for states*. By focusing on states, however, the law risks missing the major causes and consequences of IO harms. IOs often (if not mostly) originate with non-state actors and almost always affect non-state groups and, of course, individuals. To be sure, this disconnect does not disable the ability of international law to regulate entirely. International human rights law serves as a prominent example of international law's capacity to make individuals a subject of the law, as well as to require states to address non-state conduct. And, as we have discussed above, in doing so it offers protections from many harms IOs threaten. Nonetheless, the fact remains that the law operates indirectly at best (i.e., by imposing on states, rather than the actual non-state actors behind IOs, positive obligations to prevent, stop, or redress those operations occurring in or through its territory).

Ultimately, therefore, the efficacy of international law over IOs remains tethered to state behaviors. IOs carried out by non-state actors must be tied to a state for international law's negative duties to operate, while positive duties regulate states' own failure to exercise due diligence in responding to third-party IOs. Evaluating the location—and strength—of the ties between a state and a non-state actor with respect to an IO is challenging—not to mention the difficulties noted above in observing the convoluted mechanisms by which an IO takes effect in non-state actor settings. Some of these difficulties are not new: qualifying the types of assistance states accord to non-state actors has been the subject of debate for many years.[260] The standards of proof required to show legal attribution have long been under-developed just as some would say the actual legal thresholds for associating non-state actor behavior with a state are.[261] Such issues are, however, exacerbated in the IO context where so many

---

259. *See* Lanovoy, *supra* note 110, at 20–21.

260. *Compare* Vladislav Lanovoy, *The Use of Force by Non-State Actors and the Limits of Attribution of Conduct*, 28 EURO. J. INT'L L. 563 (2017), *with* Ilias Plakokefalos, *The Use of Force by Non-State Actors and the Limits of Attribution of Conduct: A Reply to Vladyslav Lanovo*, 28 EURO. J. INT'L L. 587 (2017).

261. *See* Eichensehr, *supra* note 51, at 559, 563–65 (about standards of proof and legal thresholds respectively).

harms occur not in some single, easily-observed action, but often only through small, otherwise innocuous acts or interventions.[262]

As such, existing standards are insufficient to cope with the complexity of IOs. Moreover, given the secretive nature of most IOs, applications of the rules on state responsibility are likely to produce contested outcomes as accused states deny or redirect the inquiry.[263] Further efforts are needed to iron out what "complete dependence" for the purposes of *de facto* organ status means when it comes to non-state actors orchestrating an IO campaign. Similarly, states and other stakeholders need to do more to delineate what "effective control" over a specific IO looks like.[264]

More fundamentally, it is not clear that clarifying existing standards will prove an adequate response to the challenges of IOs. Standards like "dependence" and "control" continue to assume a statist orientation for IOs that may not align with the reality of how states themselves pay "influencers" or pundits to advance their IOs and other uses of non-state actors to further their interests.[265] Sophisticated states can leverage existing legal lines, whether they are drawn broadly or precisely, to engineer IOs for which the law will not ascribe responsibility. States could, for example, associate with non-state actors in ways just less than "effective control" that encourage—and actually generate—harmful IOs. For example, non-state actor IOs that receive funding from, but are outside the direct control of, a state may evade key international legal restraints that would apply had the state engaged in instruction or directions that would trigger state responsibility.[266]

C. *Too Complex of a "Regime Complex"?*

Notwithstanding the problems with the existing law's application and orientation, we should be clear: existing international law applies

---

262. *See, e.g.*, William C. Banks, *The Bumpy Road to a Meaningful International Law of Cyber Attribution*, 113 AJIL UNBOUND 191, 192 (2019).

263. *See* Hollis & Finnemore, *supra* note 50, at 971.

264. For an application of these standards to the relationships between a state and a non-state group in the kinetic sphere, see Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosn. & Herz./Serb. & Montenegro), Merits, 2007 I.C.J. 243 (Feb. 26).

265. *See, e.g.*, Rory Sullivan, *Influencers offered money by 'Russian-linked' PR agency to discredit Pfizer vaccine*, GUARDIAN (May 25, 2021), https://www.theguardian.com/media/2021/may/25/influencers-say-russia-linked-pr-agency-asked-them-to-disparage-pfizer-vaccine [https://perma.cc/85TY-VVHT] (archived Sept. 8, 2022); Yuliya Talmazan, *Russia's media propaganda turns to 'spine-chilling rhetoric' to intimidate the West*, NBC NEWS (May 14, 2022), https://www.nbcnews.com/news/world/russia-tv-jokes-nuclear-missiles-london-putin-propaganda-ukraine-war-rcna28067 [https://perma.cc/W8N7-AJXR] (archived Sept. 8, 2022).

266. *See* ARSIWA, *supra* note 56, art. 8; Nicar. v. U.S., 1986 I.C.J. ¶¶ 80, 115.

to and regulates IOs. This fact, however, introduces a third challenge to the law's efficacy—its complexity. We have already highlighted issues of identification, assessment, and attribution that complicate the law's application in this space. But even where international lawyers can overcome these challenges, they must examine the multiple international legal obligations that each and every IO implicates.

As we have explained, states that would conduct an IO must assess its conformity with international human rights law, non-intervention, and sovereignty, to say nothing of other, more specialized, international legal regimes to which time and space precluded attention (e.g., the prohibition on the use of force, international humanitarian law, and the right to self-determination).[267] A separate set of obligations, including international human rights law and the *Corfu Channel* and no-harm principles, await where states look to address the impact of IOs by other states or non-state actors.[268] In other words, there is a very complex regime in place for IOs.[269] The question is whether this very complexity undermines the capacity of each set of rules to work together coherently to redress the threats IOs pose.

All this complexity does not end with the number of legal rules an IO may implicate. Ambiguity often lurks at the edges of rules, and sometimes at their very core. Questions of the extraterritorial reach of human rights treaty obligations are more emblematic than exceptional in this respect. And even if the relevant rules can be specified with some degree of certainty, this does not solve the conundrum of interaction between specific rules and regimes. International law's conflict avoidance tools only go so far as the text of the conflicting rules reasonably allows. And when rules are irreconcilable, there is no clear

---

267. *See supra* note 49 (explaining the authors' rationale in focusing on specific areas of international law notwithstanding open questions about how other rules like the prohibition on the use of force and self-determination operate in online contexts). Moreover, where IOs use specific methods or spaces (e.g., telecommunications or outer space), even more international legal regimes may come into play. *See, e.g.*, Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies art. III, Jan. 27, 1967, 18 U.S.T. 2410; Constitution of the International Telecommunications Union annex at 1003, art. 45(1), July 1, 1994, 1825 U.N.T.S. 331.

268. Here too, there are legal areas we have not addressed, such as the legal obligations associated with countermeasures. *See* ARSIWA, *supra* note 56, ch. II.

269. A "regime complex" refers to an overlapping set of individual regimes governing a particular issue area. *See* Kal Raustiala & David G. Victor, *The Regime Complex for Plant Genetic Resources*, 58 INT'L ORG. 277 (2004); *see also* JOSEPH S. NYE, GLOBAL COMM'N ON INTERNET GOVERNANCE, THE REGIME COMPLEX FOR MANAGING GLOBAL CYBER ACTIVITIES NO. 1 7 (May 20, 2014), https://www.cigionline.org/sites/default/files/gcig_paper_no1.pdf [https://perma.cc/5W8F-7EQZ] (archived Sept. 20, 2022).

hierarchy between conflict resolution tools.[270] When rules overlap but point in different directions (e.g., the obligation to respect freedom of expression and the *Corfu Channel* principle requiring states to protect other states from harmful IOs emanating from their territory), the outcome is uncertain. At present, international law leaves the resolution of such overlaps to interpretative techniques that may yield arbitrary results. It is difficult for international lawyers to process all these issues, let alone on the short timelines that the risks posed by IOs may require.

D. *The Absence of Effective Enforcement for Internationally Wrongful IOs*

Notwithstanding the challenges of application, orientation, and complexity, there are areas of international law that offer real promise when it comes to regulating IOs. International human rights law, for example, encompasses an extensive set of rights (to life, health, political participation, freedom of expression, freedom of thought, privacy, etc.) that both prohibit states from engaging in violative IOs and require states to ensure those within their jurisdiction are protected from third-party violations of these rights. These laws clearly govern what states do internally and, depending on the appropriate approach to extraterritoriality, may also restrict state interference with the human rights of people abroad.

Enforcement of these rights, however, is often a challenge. For starters, there is an absence of legal rhetoric in the rising use of accusations about IOs.[271] It is notable, for example, that despite public attribution—and complaints—by the United States that Russia had conducted a series of IOs targeting the 2016 US presidential election, the United States never publicly claimed that Russia violated international law in doing so.[272] And this represents the norm rather than the exception as states and others increasingly identify IOs in the wild. There may be multiple explanations for why states have done little to invoke international law in the IO context. For instance, they may be unclear on the contours and contents of the law given the application challenges we have described above.[273] States may be silent as a matter of political expediency. Or they may see silence as a way to preserve operational flexibility (i.e., to respond to IOs with IOs rather

---

270. *See, e.g.*, Vienna Convention on the Law of Treaties art. 30, Jan. 27, 1980, 1155 U.N.T.S. 331; Martti Koskenniemi, Rep. of the Study Group of the Int'l L. Comm'n, *Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law*, 47 *et seq.*, U.N. Doc. A/CN.4/L.682 (Apr. 13, 2006).

271. *See* Finnemore & Hollis, *supra* note 50, at 971; Eichensehr, *supra* note 51, at 529.

272. *See* Finnemore & Hollis, *supra* note 50, at 982.

273. Hollis, *Fifth Report*, *supra* note 250, at 7, 13.

than rely on the law to forestall or deter such conduct).[274] Alternatively, states may find greater appeal in other regulatory tools besides international law. A number of states have used domestic criminal laws—and indictments of individuals under them—to address harmful IOs.[275] Further, states may decide to let the online platforms where IOs occur take the lead in devising and enforcing the rules of the road for IOs, independent of international law. (For their part, social media and other technology companies have begun to call for state regulation of their platforms, even as they turn to international law as a source of authority for their decisions on what content to keep up or take down.[276]) Whatever the reasons, if international law is not invoked, international law cannot be enforced.

Second, international law's enforcement mechanisms are relatively limited. Certainly, some human rights enforcement can—and regularly does—occur. Regional human rights treaties in Europe and the Americas encompass international courts—the ECtHR and the Inter-American Court of Human Rights—that can identify human rights violations by member states and direct remedial measures in response.[277] There is, moreover, nothing to preclude international tribunals from exercising jurisdiction over a case involving IOs that implicate rights or obligations provided for in their constituent treaty or jurisdictional instrument. Of course, member states may resist implementing particular judgements, but that is not a problem confined to IOs.

---

274. This possibility has already occurred in the cyber context. Iran, for example, reportedly never challenged the alleged US and Israeli role in Stuxnet as a use of force or even an armed attack (triggering a right of self-defense), preferring instead to deploy its own cyber-operations against US financial targets without any legal framing at all. *See, e.g.*, Mike Mount, *U.S. Officials believe Iran behind recent cyber attacks*, CNN (Oct. 16, 2012), https://www.cnn.com/2012/10/15/world/iran-cyber/index.html [https://perma.cc/A734-UYB2] (archived Sept. 8, 2022). Several Iranians were later indicted for their participation in these operations. *See Seven Iranians Working for Islamic Revolutionary Guard Corps-Affiliated Entities Charged for Conducting Coordinated Campaign of Cyber Attacks Against U.S. Financial Sector*, DEP'T. OF JUST. (Mar. 24, 2016), https://www.justice.gov/opa/pr/seven-iranians-working-islamic-revolutionary-guard-corps-affiliated-entities-charged [https://perma.cc/QV35-C3Z3] (archived Sept. 20, 2022).

275. *See* Eichensehr, *supra* note 51, at 532.

276. *See, e.g.*, Lauren Feiner, *Microsoft's Brad Smith says tech regulation is coming, so industry should participate in shaping it*, CNBC (April 13, 2022), https://www.cnbc.com/2022/04/13/microsoft-president-brad-smith-tech-industry-regulation-coming.html [https://perma.cc/6KP2-HF63] (archived Sept. 20, 2022); Alex Warofka, *Human Rights Impact of Facebook in Myanmar*, FACEBOOK (Nov. 5, 2018), https://about.fb.com/news/2018/11/myanmar-hria [https://perma.cc/M2FC-NVXU] (archived Sept. 8, 2022).

277. In addition to international courts, various human rights treaty bodies may have—or at least assert—jurisdiction over the operation of the rights accorded by the treaty that created them. But absent consent to specific dispute settlement procedures, most of these bodies lack enforcement tools to enforce their own opinions vis-à-vis states parties.

Outside of these regional treaties, however, human rights enforcement becomes more daunting, to say nothing of enforcement of obligations like non-intervention or the *Corfu Channel* principle. Some treaties do contain specific compliance mechanisms, but few of these offer obvious opportunities for enforcing international law over IOs.[278]

IOs that constitute internationally wrongful conduct do entitle other affected states to engage in acts of retorsion or countermeasures. The difficulty is that acts of retorsion have yet to demonstrate much efficacy or accountability for IOs carried out by states or non-state actors.[279] At the same time, there has been robust discussion for the last several years of the potential for states to employ countermeasures—otherwise unlawful measures, "the wrongfulness" of which is "precluded" when adopted as a response to a prior internationally wrongful act.[280] Like legal accusations, however, the practice of states employing countermeasures (at least openly) is largely non-existent. This may be due to the strict conditions under which international law permits countermeasures. Indeed, even where a state is entitled to take countermeasures, it can only do so (i) after calling upon the responsible state to fulfil its obligations;[281] (ii) upon notifying and offering to negotiate with said state, except in cases of urgency;[282] (ii) for an appropriate purpose;[283] (iii) via non-escalatory and (if possible) reversible means;[284] (iv) proportionally;[285] and (v) with respect for fundamental human rights.[286] If a state invokes countermeasures inappropriately (e.g., it is in error with respect to the international wrongfulness of the IO to which it responds), the state's behavior is itself wrongful. And even if the IO triggering countermeasures did violate international law, states are still responsible for any harms to third parties their countermeasures

278. Of course, just because a treaty has designated authorities that can enforce compliance does not mean they will actually be used—as witnessed by the repeated blocking of UN Security Council action by permanent member vetoes. *See* U.N. Charter ch. VII.

279. Acts of retorsion consist of unfriendly, but otherwise lawful behavior. *See* ARSIWA, *supra* note 56, cmt. on ch. II, at 128.

280. *See, e.g.*, Michael N. Schmidt, *'Below the Threshold' Cyber Operations: The Countermeasures Response Option and International Law*, 54 VA. J. INT'L L. 697 (2013).

281. *See* ARSIWA, *supra* note 56, art. 52(1)(a).

282. *See id.* arts. 52(1)(b), 52(2).

283. *See id.* art. 49; TALLINN MANUAL 2.0 *supra* note 14, at 116.

284. *See* ARSIWA, *supra* note 56, art. 49.

285. *See id.* art. 51. This requires the state to consider the gravity of the internationally wrongful act but does not require that the countermeasure take place in the same domain as the original internationally wrongful act (i.e., a cyber act does not require a cyber response).

286. *See id.* art. 50. The *Articles on State Responsibility* also expressly state that countermeasures must not conflict with the prohibition on the use of force, obligations of a humanitarian character prohibiting reprisals, and preemptory norms of international law; they must also respect the inviolability afforded missions and their personnel under diplomatic and consular law. *Id.*

cause. Thus, states must limit any of their countermeasures' unlawful impacts to the targeted state, a requirement that may be difficult to satisfy in the cognitive context. Several states have endorsed a right to deploy countermeasures collectively—either in response to breaches of *erga omnes* and *erga omnes partes* obligations or at the request of the injured state.[287] However, other states and the *Tallinn Manual 2.0* contest the legality of such moves, insisting that only the state injured by an internationally wrongful act may do so.[288] All these conditions and debates may have denuded countermeasures of much practical utility in enforcing international law in the ICT context, including with respect to wrongful IOs.

Many of the foregoing enforcement problems are not unique to IOs—they persist in international relations generally, or at least in the ICT context. Yet the nature (and novelty) of IOs in the digital environment may exacerbate these enforcement challenges. Although some IOs may come directly from the state (e.g., an IO by a government targeting its own population), most will occur in online fora owned and operated by non-state actors (e.g., social media and other technology companies). It is these companies that constitute the first line of defense against wrongful IOs as well as the most likely to take responsive measures that clash with the conditions under which international human rights law permits limitations on speech. Identifying and assessing IOs will inevitably require cooperation with these platforms; cooperation that may be complicated by states' own diverse approaches to IO issues. A social media company like Facebook may look to international law to delineate its terms of service or content moderation guidelines.[289] At the same time, it must navigate dozens of domestic or supranational legal regimes, which may impose requirements, prohibitions, and permissions that could compete, if not conflict, with international law. In such circumstances, enforcing international law becomes an exercise in working through corporations

---

287. *See, e.g.*, Kersti Kaljulaid, President of Estonia, Opening at CyCon 2019 (May 29, 2019), https://www.president.ee/en/official-duties/speeches/15241-president-of-the-republic-at-the-opening-of-cycon-2019/index.html [https://perma.cc/82VP-UQLD] (archived Sept. 8, 2022); New Zealand Statement, *supra* note 186, ¶ 22 (2020) (stating that the country is "open to the proposition that victim states, in limited circumstances, may request assistance from other states in applying proportionate countermeasures to induce compliance by the state acting in breach of international law").

288. *See* French Position on International Law Applications to Cyberspace, *supra* note 212, at 4; TALLINN MANUAL 2.0 *supra* note 14, at 130 and accompanying commentary. The *Draft Articles of State Responsibility* expressed uncertainty over this question and declined to offer a resolution. ARSIWA, *supra* note 56, cmt. to art. 54, at 139.

289. *See, e.g.*, FACEBOOK OVERSIGHT BD., OSB CASE DECISION 2021-001-FB-FBR (2021), https://www.oversightboard.com/sr/decision/2021/001/pdf-english [https://perma.cc/HR2A-2DGJ] (archived Sept. 8, 2022) (about Donald Trump).

as intermediaries, a pathway that is often alien to states accustomed to more direct roles in international law enforcement.

Finally, in discussing international law, it is important not to lose sight of the current conditions. Simply put, the world is awash in IOs, many of which threaten real harms to life, health, political processes, and even the very sovereignty of states. In such circumstances, the real enforcement challenge lies not just in achieving compliance with whatever conduct international law currently covers but also in mitigating IO harms that fall outside its ambit. After all, international law exists not for its own sake but to protect fundamental values and interests of states and their peoples. If the law fails to accord those protections under existing provisions (and enforcement mechanisms), the problem may lie as much with the state of the law as with the IOs that it permits.

Taken together, international law's *enforcement* mechanisms—the limited remedies available to respond to internationally wrongful acts—have yet to demonstrate a compliance pull that can adequately forestall harmful IOs from occurring or otherwise remediate their harmful effects on states, corporations, and individuals. Indeed, the challenges with existing law could actually be incentivizing rather than deterring the use of IOs by state and non-state actors alike.

V. DO WE NEED AN INTERNATIONAL LAW FOR INFORMATION OPERATIONS (ILIO)?

As the preceding discussion shows, states and other stakeholders face extensive challenges in using existing international law to regulate IOs in the digital age. The existing system has issues of application, orientation, complexity, and enforcement that may encourage, rather than deter, many of the harmful IOs it should forestall. And although many IOs undoubtedly remain *sub rosa* there is increasing evidence of their proliferation.

Fifteen years ago, one of us wrote an article calling for states to devise an international law for information operations—or ILIO.[290] At the time, the IO definition conflated IOs with cyber operations more broadly. And in the ensuing years, states and scholars have debated the need for—and utility of—devising specific legal regimes for cyber operations.[291] Such calls help explain the nascent push to devise a UN

290. *See generally* Duncan B. Hollis, *Why States Need an International Law for Information Operations*, 11 LEWIS & CLARK L. REV. 1023 (2007).

291. *See generally* Oona Hathaway, Rebecca Crootof, Philip Levitz, Haley Nix, Aileen Nowlan, William Perdue & Julia Spiegel, *The Law of Cyber-Attack*, 100 CALIF. L. REV. 817 (2012); JACK GOLDSMITH, HOOVER INST., CYBERSECURITY TREATIES: A SKEPTICAL VIEW (2011), https://www.hoover.org/sites/default/files/research/docs/futurechallenges_goldsmith.pdf [https://perma.cc/YTP8-C4EU] (archived Sept. 20, 2022); TALLINN MANUAL 2.0, *supra* note 14, at 2–3.

Convention on Cybercrime and ongoing debates over the need for other cybersecurity treaties, whether a Digital Geneva Convention or one that simply codifies the existing rules.[292]

In contrast to the attention cyber operations have received, questions about international legal regulation of IOs have been surprisingly sparse, especially considering the prominent ways IOs have impacted international relations and our individual lives of late. In light of the rising significance of IOs, we believe it is time for states and other stakeholders to weigh the benefits (and costs) of devising international law obligations specific to IOs.

It is not, however, our ambition to claim the world needs such obligations. Neither is it to reject an ILIO outright. Our aim is more modest: to catalyze states to weigh the pros and cons of pursuing an ILIO project itself.

Fostering an ILIO would undoubtedly have multiple advantages that could address the challenges faced by the existing law. For starters, it could reaffirm the general obligations under international law reviewed above, while also elaborating or clarifying their contents in the specific IO context. An ILIO could, for example, resolve questions over the meaning of "coercion" online. It could clarify that the no-harm principle includes physical and non-physical harms to life, health, and other rights caused by IOs emanating from a state's territory or jurisdiction. It could also address the extraterritorial application of those rights, without needing to depend on the (unlikely) resolution of existing debates over the extraterritorial application of human rights more generally.

Clarifying the application of international law to IOs need not, however, be limited to elaborations of existing law. It would present an opportunity to design obligations tailored to the nature of IOs and the harms they pose. An ILIO might, for example, explicitly prohibit certain types of "domestic IOs," where state officials spread mis-, dis-, or malinformation about the state's own electoral processes or other matters involving the basic rights of their citizens (e.g., pandemic response). It might prohibit IOs that cause serious adverse consequences to the conduct of an electoral process in another state. Or it could prohibit IOs that incite more violations of international law than the current rules reach (e.g., going beyond incitement to genocide,

---

292. *See, e.g.*, Katitiza Rodriguez & Meri Baghdasaryan, *UN Committee To Begin Negotiating New Cybercrime Treaty Amid Disagreement Among States Over Its Scope*, ELEC. FRONTIER FOUND. (Feb. 15, 2022), https://www.eff.org/deeplinks/2022/02/un-committee-begin-negotiating-new-cybercrime-treaty-amid-disagreement-among [https://perma.cc/3AHB-EAPV] (archived Sept. 8, 2022); Brad Smith, *The need for a Digital Geneva Convention*, MICROSOFT ON THE ISSUES (Feb. 14, 2017), https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/ [https://perma.cc/V4Q8-WU94] (archived Sept. 8, 2022).

discrimination, hostility, violence, and war to cover incitement to other acts incompatible with the internal order or the security of a territory of a state, or false statements that "harm good international understanding"[293]). This could be done by, for example, reviving and building on the 1936 Convention concerning the Use of Broadcasting in the Cause of Peace.[294] Constructing an ILIO via a treaty would avoid problems inherent in the development of customary international law. In particular, the secretive nature of IOs raises questions about when IOs constitute "state practice" and what legal salience attaches to state responses to them.[295]

Similarly, an ILIO could reorient international law in ways tailored to the threats IOs pose. IOs could be defined and those whose harms warrant international legal regulation could be delimited. It would provide opportunities to move past the existing law's statist orientation and emphasize the threats posed by non-state actor IOs. An ILIO could, for example, offer a specific prohibition of IOs involving types of terrorist or extremist content that are generally agreed upon by states via multilateral or bilateral treaties, unilateral government statements, or non-binding collective documents, such as the Christchurch Call.[296] It could address the "cognitive" core of IOs by laying out clear rules on causation and proposing collective processes for identifying the emergence and effects of IOs, online and offline. Attribution issues could be addressed by delineating standards of proof and/or the requisite linkage required to establish state responsibility in this context. The latter move could significantly impact the law's efficacy if it recalibrates state involvement in non-state actor IOs. This could potentially reduce the ongoing widespread use of non-state actor proxies, without the significant burden of proving dependence or control that triggers state responsibility currently.

By devising an ILIO, states would also address the complexity problem. A tailor-made legal regime could take a *lex specialis* form. This would allow states and their lawyers to have a single reference point for what IOs they can (and cannot) pursue as well as the menu

---

293. 1936 Broadcasting Convention, *supra* note 240, art. 3.

294. *See id.* The Convention was recognized by the UN General Assembly in 1954 as "an important element in the field of freedom of information." G.A. Res. 841 (IX), pmbl. ¶ 1 (Dec. 17, 1954). On the same occasion, the General Assembly set in motion a plan to update and supplement the Convention by means of drafting a new Protocol. *See id.* ¶ 2. However, Cold War divisions stalled the project, which was subsequently abandoned. *See* MICHAEL G. KEARNEY, THE PROHIBITION OF PROPAGANDA FOR WAR IN INTERNATIONAL LAW 28–33 (2007).

295. In addition, relying on custom to develop around IOs risks giving priority to state practices by the states who get caught—i.e., whose IO is disclosed publicly—which may not be the states who international law needs to regulate. It would be better to have high-capacity actors buy-in from the outset as would be the case if states negotiate and conclude an ILIO treaty.

296. *See* Hollis, *Why States Need an International Law for Information Operations*, *supra* note 290, at 1059–61.

of positive steps the law would expect of them when harmful IOs by others occur.

Finally, the benefits of an ILIO could extend to enforcement. The mere act of formally tailoring international law obligations to IOs could have important signaling effects for states, demonstrating a shift in expectations for law-abiding states that could in turn deter or prevent questionable IOs states might otherwise pursue. It would also open up opportunities for legal protests that states appear reluctant to engage in under existing, general international law rules. Moreover, if states adopted an ILIO in treaty form, they could incorporate treaty-specific enforcement mechanisms to investigate, identify, assess, apply, and respond to covered IOs. Tailor-made enforcement measures could carefully balance the risks to human rights while simplifying the opportunities for enforcement compared to the unwieldy world of countermeasures today. Finally, an ILIO could be accompanied by the establishment of a specific treaty body, charged with mandatory or optional adjudication of disputes at least between states parties.

In listing the positive potential of an ILIO, we do not mean to ignore the risks or costs it would entail. Treaty making has a long history of sucking up extensive time and resources states could put to other purposes. Debates over the format and forum for an ILIO would be extensive and contentious. Existing differences in the UN Open Ended Working Group on developments in the field of information and telecommunications in the context of international security were exacerbated by questions of multistakeholder participation[297]—a topic that would have equal importance in the IO context given the aforementioned central role played by online platforms and other technology firms. Similar fights would likely arise over the form an ILIO should take; differences will inevitably emerge over whether to formally devise rules (i.e., a treaty) or pursue soft law.

The establishment of a *lex specialis* regime would raise new—and complex—questions over its relationship with existing rules under treaties and customary international law. Would a new regime displace existing rules altogether? Or would they apply concurrently? How would it affect, if at all, the interpretation of existing rules, including existing efforts to elaborate how international law applies to cyber operations more generally? This is especially relevant if only some states sign up to such an initiative. States should diligently avoid any prejudice to the scope of protections under the current international legal framework; the goal of a new regime should be to offer additional, tailor-made protection, not to erode existing legal guarantees.

---

297. *See Modalities of multistakeholder participation*, DIGIWATCH (Dec. 13, 2021), https://dig.watch/event/un-oewg-2021-2025-1st-substantive-session/modalities-of-multistakeholder-participation [https://perma.cc/CW76-QPEG] (archived Sept. 8, 2022) (providing a report on the U.N. Open-ended Working Group's first substantive session).

Additionally, recent treaty efforts provide a stark reminder that success is not guaranteed; negotiations of an ILIO could fail. Even if successful, a tailor-made ILIO might not achieve the necessary adoption, incentivizing free riders or, worse, making the whole endeavor ineffective.

An additional complication for an ILIO is that ICTs are constantly evolving. This risks desuetude for an ILIO if its regulatory framework became technically obsolescent. Regulating IOs specifically, including the tools and techniques by which it occurs, would also risk spill-over effects. It could instantiate technologies that themselves negatively impact—or even violate—human rights. International lawyers should be careful, therefore, to avoid offering a cure worse than the disease. Having a special regime for IOs that could potentially include its own enforcement regimes could lead to the unwillingness of other tribunals and mechanisms to engage with these questions. Conversely, it could generate opportunities for fragmentation of both norms and institutions. Rather than having two alternatives—IOs covered under both a special regime and the general one—we may risk a carveout from the general regime as well.

VI. CONCLUSION

Today, IOs are being deployed by states and non-state actors alike for a range of purposes that implicate a host of individual, corporate, and state interests while exacerbating inequalities in international relations. Our Article has examined how international law deals with this phenomenon, particularly its increasingly prominent online forms. We examined how international law currently regulates IOs from the perspective of *lex lata* and the opportunities that may reside in developing new law specifically tailored to IOs as a matter of *lex ferenda*. We provided an overview of existing international legal rules applicable to IOs by analyzing obligations arising under international human rights law, the principles of non-intervention and sovereignty, and the *Corfu Channel* and no-harm principles. We explored the effectiveness of the current regime through four lenses: (i) applicability, (ii) orientation, (iii) complexity, and (iv) enforcement. Given the range of identified shortcomings, we inquired into the desirability of establishing new law on IOs—an ILIO.

That IOs can produce harms for individuals, groups, and states is beyond contention. Information has, for centuries, been twisted, manipulated, and weaponized for personal or political gain. In today's digital environment, IOs can spread at an unprecedented speed, reach any locality, and foster exclusion through the siloization of communities. One of the most apt—and tragic—demonstrations of the risks of information silos is on display in Russia, where many citizens continue to trust the Kremlin's rhetoric on the reasons, nature, and

success of their so-called "special military operation" in Ukraine.[298] But for the blind support of many Russians, manipulated for years through relentless domestic IOs, this war of aggression could have ended soon after its initiation.

The risks IOs pose are clear. How to address those risks is the sharp-ended question facing the international community right now. Undoubtedly, a key part of any answer lies in international law. States, international organizations, and other actors have shown a firm commitment to international law as a necessary framework for stable and predictable interactions online and offline.[299] And there is no doubt that existing international law contains a wide range of rules that already regulate the ways in which states, non-state actors, and individuals carry out and engage with IOs. While these rules provide important protections against harmful IOs, they are not, however, regularly *tailored to* such operations, especially those occurring online. International law is instead often characterized by a generic and state-centric orientation. Perhaps partly due to these application and orientation deficiencies, international law suffers from under-enforcement, which in turn minimizes its deterrent effect.

To be clear, we are not suggesting that the existing legal framework does not regulate IOs nor that it is fundamentally ill-equipped to address its risks. Rather, our goal is to reorient the conversation around international law's *efficacy*—how well it operates today and explanations for the challenges it faces. To facilitate that conversation, we have called for more attention to the benefits (and risks) inherent in developing a *lex specialis* for IOs. Even if a *lex specialis* for IOs never sees the light of day, it is a discussion that can incentivize states and other stakeholders to clarify their positions on the content of existing law. It is also possible that the specific dangers of IOs will make relevant actors more likely to reach agreement on specific protections for IOs—as was the case with the 1936 Broadcasting Convention—even if they are averse to expanding international law protections more generally. The seeds of such dynamics can be seen in discussions on the legality of third-party or collective countermeasures. Indeed, there are signs that the particular challenges of the ICT environment have made such measures more amenable to states, arguably going in the direction of creating a *lex specialis* for cyber countermeasures. Specialized regimes under international law also often come with their own institutional

298. *See* Valerie Hopkins, *Ukrainians Find That Relatives in Russia Don't Believe It's a War*, N.Y. TIMES (March 6, 2022), https://www.nytimes.com/2022/03/06/world/europe/ukraine-russia-families.html [https://perma.cc/TJV4-6V9D] (archived Sept. 20, 2022).

299. *See* U.N. Open-Ended Working Group On Developments In The Field Of Information And Telecommunications In The Context Of International Security, *Final Substantive Report* ¶ 34, U.N. Doc. A/AC.290/2021/CRP.2 (Mar. 10, 2021).

structure, including mechanisms for review. International law, with its lack of a system of courts with general compulsory jurisdiction, might thus benefit from new review mechanisms vis-à-vis IOs, as such mechanisms could exert a compliance pull, clarify the interpretation of rules, and operationalize international legal protections for states and individuals alike.

Whichever direction the international community decides to take, effective protection against harmful IOs will require legal clarity and avenues for effective enforcement. There is no doubt that international law is a viable and necessary tool in countering the risks posed by IOs. It is high time for states and other relevant stakeholders to recognize—and unlock—its potential to regulate IOs in a way that places human and societal interests front and center.

Vanderbilt Journal
of Transnational Law

The *Vanderbilt Journal of Transnational Law* (*Journal*) (USPS 128-610) is published five times a year (Jan., Mar., May, Oct., Nov.) as part of the International Legal Studies Program by the Vanderbilt University Law School, 131 21st Avenue South, Room 152, Nashville, TN 37203. The *Journal* examines legal events and trends that transcend national boundaries. Since its foundation in 1967, the *Journal* has published numerous articles by eminent legal scholars in the fields of public and private international law, admiralty law, comparative law, and domestic law of transnational significance. Designed to serve the interests of both the practitioner and the theoretician, the *Journal* is distributed worldwide.

The preferred and most efficient means of submission is through Scholastica at https://scholasticahq.com. However, other modes of submission are accepted in print or by e-mail attachment.

Footnotes must conform with *The Bluebook: A Uniform System of Citation* (most recent edition), and authors should be prepared to supply any cited sources upon request. Authors must include a direct e-mail address and phone number at which they can be reached throughout the review period.

Subscriptions beginning with Volume 49 are $35.00 per year (domestic), $40.00 per year (foreign); individual issues are $10.00 domestic and $11.00 foreign. Orders for subscriptions or single issues may enclose payment or request billing and should include the subscriber's complete mailing address. Subscriptions will be renewed automatically unless notification to the contrary is received by the *Journal*. Orders for issues from volumes prior to and including Volume 16 should be addressed to: William S. Hein & Co., Inc., 2350 North Forest Road, Getzville, NY, 14068.

Please send all inquiries relating to subscriptions, advertising, or publication to: Program Coordinator, Vanderbilt Journal of Transnational Law, Vanderbilt Law School, 131 21st Avenue South, Room 152A, Nashville, Tennessee, 37203, Phone: (615) 322-2284, Facsimile: (615) 322-2354, Email Address: faye.johnson@law.vanderbilt.edu.

Class "Periodicals" postage is paid at Nashville, Tennessee, and additional mailing offices. POSTMASTER: Send address changes to Program Coordinator, Vanderbilt Journal of Transnational Law, Vanderbilt Law School, 131 21st Avenue South, Room 152A, Nashville, Tennessee, 37203.

The *Journal* is indexed in *Contents of Current Legal Periodicals*, *Current Law Index*, *Index to Legal Periodicals*, and *Index to Foreign Legal Periodicals*.

Antidiscrimination Policy: The *Journal of Transnational Law* abides by the Vanderbilt University Equal Opportunity Policy, available at http://www.vanderbilt.edu/student_handbook/university-policies-regulations/#equal-opportunity. The viewpoints expressed by authors do not necessarily represent the views of Vanderbilt University Law School.

Cite as: VAND. J. TRANSNAT'L L.

***