PennState Dickinson Law

DICKINSON LAW REVIEW
PUBLISHED SINCE 1897

Volume 126 | Issue 1

Fall 2021

Taking Exception to Assessments of American Exceptionalism: Why the United States Isn't Such an Outlier on Free Speech

Evelyn Mary Aswad
University of Oklahoma College of Law

Taking Exception to Assessments of American Exceptionalism: Why the United States Isn't Such an Outlier on Free Speech

Evelyn Mary Aswad*

ABSTRACT

One of the most significant challenges to human freedom in the digital age involves the sheer power of private companies over speech and the fact that power is untethered to existing free speech principles. Heated debates are ongoing about what standards social media companies should adopt to regulate speech on their platforms. Some have argued that global social media companies, such as Facebook and Twitter, should align their speech codes with the international human rights law standards of the United Nations ("U.N."). Others have countered that U.S.-based companies should apply First Amendment standards. Much of this debate is premised on a fundamental misunderstanding about the scope of freedom of expression protections under U.N. standards.

This Article addresses that pervasive misunderstanding by engaging in a detailed comparison of key doctrines underlying both bodies of law. The Article provides the first in-depth comparison of U.S. and U.N. standards on freedom of expression since the U.N. human rights machinery adopted pivotal interpre-

* The author is the Herman G. Kaiser Chair in International Law and the Director of the Center for International Business & Human Rights at the University of Oklahoma College of Law. Previously, she served as the director of the human rights law office at the U.S. State Department from 2010–2013 and was an attorney-adviser in that human rights law office from 2004–2009. This article was made possible through the support of the John S. and James L. Knight Foundation. The author would like to thank the CATO Institute for including her in programming and discussions that helped to inform her thinking with respect to this Article. The author thanks in particular former ACLU President and Professor Emerita Nadine Strossen for her comments on this Article and an ongoing dialogue about U.S. and U.N. approaches to speech. The author is grateful to Michael McConnell, Rick Tepker and Elizabeth Cassidy for their comments on draft sections of this Article. The author also thanks Cooper Eppes, Robert Rembert, and Morgan Vastag for their excellent research assistance. The views are solely those of the author.

tations of this human right a decade ago. The Article finds that both standards provide a principled and disciplined approach to speech restrictions by creating a presumption in favor of speech, prohibiting unduly vague and overbroad speech restrictions, mandating that only narrowly tailored burdens on speech be authorized, and requiring that any restrictions serve important public interest objectives.

While this Article does not argue that the two bodies of law completely converge, it does maintain that the key doctrines they share should inform—and perhaps transform—the ongoing debate about what standards social media companies should use in curating content on their platforms. U.N. standards are more protective of speech than is generally understood to be the case and provide a framework that can be translated to the context of private sector content moderation.

TABLE OF CONTENTS

INTRODUCTION

Many scholars have opined that the United States is an outlier on freedom of expression when comparing the breadth of First Amendment protections with the regulation of speech in other countries.[1] The legal academy has generally extended this outlier assessment to encompass the view that the First Amendment is significantly—and perhaps irreconcilably—more protective of speech than the United Nations' ("U.N.") international human rights law standard.[2] But, recently, some prominent U.S. free speech experts

1. *See, e.g.*, Harold Hongju Koh, *On American Exceptionalism*, 55 STAN. L. REV. 1479, 1483 (2003) (observing that "the U.S. First Amendment is far more protective [of speech] than other countries' laws of hate speech, libel, commercial speech, and publication of national security information"); ANN-MARIE SLAUGHTER, A NEW WORLD ORDER 243 (2004) (noting that "if the judges of the U.S. Supreme Court thought that they were playing to a global as well as a national audience, they might readily acknowledge that U.S. First Amendment jurisprudence is on the extreme end of the global spectrum for protecting speech"); Ronald J. Krotoszynski, *Free Speech Paternalism and Free Speech Exceptionalism: Pervasive Distrust of Government and the Contemporary First Amendment*, 76 OHIO STATE L.J. 659, 659 (2015) (describing the U.S. approach to free speech as "a global anomaly"); Scott H. Greenfield, *The Internet and Free Speech Calculus*, SIMPLE JUST. (Oct. 7, 2012), https://bit.ly/3kdafBE [https://perma.cc/X64K-QBBR] ("Compared with almost every country in the world, the United States is an outlier when it comes to free expression." (quoting Harvard Law School Professor Noah Feldman)). Practitioners have also noted that the U.S. approach to speech differs significantly even from like-minded democracies. *See, e.g.*, Peter Scheer, *The U.S. Is Alone Among Western Democracies in Protecting "Hate Speech." Chalk It Up to a Healthy Fear of Government Censorship*, FIRST AMEND. COAL. (Mar. 14, 2011), https://bit.ly/2Vq0c1T [https://perma.cc/N9EV-ENW5] (noting that "[t]he United States is an outlier when it comes to freedom of expression.").

2. *See, e.g.*, Mari J. Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320, 2348 (1989) (highlighting the "failure of American law to accept this emerging [international human rights law ban on racist hate speech] reflects a unique first amendment jurisprudence"); Dawn C. Nunziato, *How (Not) to Censor: Procedural First Amendment Values and Internet Censorship Worldwide*, 42 GEO. J. INT'L L. 1123, 1130 (2011) (determining that "[a]lthough free speech is granted some protection by international treaties, this protection is subject to a host of limitations and exceptions—far more than under the First Amendment"); Jean-Marie Kamatali, *The U.S. First Amendment Versus Freedom of Expression in Other Liberal Democracies and How Each Influenced the Development of International Law on Hate Speech*, 36 OHIO N.U. L. REV. 721, 730–31 (2010) (arguing that the international freedom of expression standard "contains restrictions going beyond those permissible under current First Amend-

(particularly the former President of the ACLU) have begun to question that long-standing view, noting the two standards may be much more similar than is generally understood to be the case.[3]

The distinction between the First Amendment's and the U.N.'s approaches to free speech is gaining attention in the context of debates about which standards U.S. social media companies should voluntarily employ to judge user-generated content on their platforms,[4] a critically important process that is known (somewhat blandly) as "content moderation."[5] As private actors, U.S. companies are not bound to respect First Amendment speech protections.[6] Similarly, international human rights law obligations

---

ment jurisprudence" and contains mandatory bans on speech that are "in plain conflict with the strictures of the U.S. Constitution"); *see also* Jack Goldsmith, *Should International Human Rights Law Trump U.S. Domestic Law?*, 1 CHI. J. INT'L. L. 327, 330–31 (2000) (assessing that international human rights protections for expression "are probably inconsistent with First Amendment free speech rights").

3. *See* NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 211 (2019) (stating that "the international human rights standard is more analogous to U.S. law's speech-protective standards concerning hateful speech than to the less speech protective standards of other countries" or those in "regional human rights treaties, including the European Convention on Human Rights"); *see also* John Samples, *International Law and "Hate Speech" Online*, CATO INST.: CATO AT LIBERTY (July 23, 2020, 2:40 PM), https://bit.ly/3ASYczp [https://perma.cc/VE63-TVYB] (commemorating the observations of a CATO Vice President that the U.N.'s international standard requires any limits on speech to meet certain thresholds, which may protect more speech than is commonly understood to be the case).

4. *See, e.g.*, Nathaniel Persily, *Live Debate with Stanford Law School Professor Nate Persily—Constitutional Free Speech Principles Can Save Social Media Companies from Themselves*, STAN. L. (Mar. 5, 2019), https://stanford.io/2UCAYNB [https://perma.cc/YNK5-3AFV] (hosting a debate among practitioners and academics about whether platforms should adopt U.S. free speech principles or international human rights norms); *Social Media, Election 2020, and Online Speech*, NAT'L CONST. CTR. (Nov. 3, 2020), https://bit.ly/3ANxN65 [https://perma.cc/FBS4-LKCS] (hosting a debate on platform governance with commentators taking differing views on the utility of First Amendment and international human rights law approaches).

5. *See* Jillian C. York & David Greene, *How to Put Covid-19 Content Moderation into Context*, BROOKINGS: TECH STREAM (May 21, 2020), https://brook.gs/3hwqR5N [https://perma.cc/HQ26-9988] (suggesting that the phrase "content moderation" is "a euphemism for what is actually private censorship–at scale"). This Article uses the phrase "content moderation" to refer to the rules and policies companies apply to judge speech on their platforms and does not encompass situations in which governments require removal of user-generated speech.

6. *See* Marvin Ammori, *The "New"* New York Times*: Free Speech Lawyering in the Age of Google and Twitter*, 127 HARV. L. REV. 2259, 2283–84 (2014) (noting that the First Amendment does not legally bind private actors but does influence technology companies' speech policies). In addition, Section 230 of the U.S. Communications Decency Act shields (with a few exceptions) online intermediaries

generally do not apply directly to corporate actors.[7] Given that social media platforms wield power over the speech of billions,[8] this legal blackhole has caused concern about private actors regulating online speech in a manner that is untethered to any principled standards.[9]

The U.N.'s top expert on freedom of expression and others have called on multinational platforms to align their content moderation with the international human rights standards on speech set forth in the U.N.'s International Covenant on Civil and Political Rights ("ICCPR").[10] This call is grounded in the U.N. Guiding Principles on Business & Human Rights ("UNGPs"), a global framework that expects corporations to respect the human rights

---

from liability for third party content, which gives companies discretion in moderating content on their platforms. *Id.* at 2286–90 (discussing 47 U.S.C. § 230 (2006)).

7. As noted by the U.N. Secretary General's Special Representative on human rights and business, there is "little authoritative basis in international law—hard, soft, or otherwise"—to conclude that existing international human rights law applies directly to corporate actors. John Ruggie (Special Representative of the Secretary General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises), *Interim Rep. of the Special Representative of the Secretary General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises*, ¶ 60, U.N. Doc. E/CN.4/2006/97 (Feb. 22, 2006).

8. For example, Google's YouTube reported that over two billion individuals use its site. *YouTube For Press*, YOUTUBE: OFF. BLOG, https://bit.ly/3AIBREG [https://perma.cc/R83G-ZF6A] (last visited Aug. 7, 2021). Facebook has 2.74 billion users around the world who check their accounts at least once a month. Heather Kelly, *Why It's Easy to Hate Facebook but Hard to Leave*, WASH. POST (Nov. 19, 2020, 7:00 AM), https://wapo.st/3qYQ7EY [https://perma.cc/H2HX-FBBT].

9. *See, e.g.*, Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1627 (2018) (observing that companies generally remove offensive speech because of "the threat that allowing such material poses to potential profits based in advertising revenue"); Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 31 (2018) (asking "how much will it matter ten or fifteen years from now that the First Amendment (and international human rights law) protect freedom of expression, if most communication happens online and is regulated by private platforms").

10. *See* David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶¶ 3, 45, 70, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018) [hereinafter *Special Rapporteur 2018 Report*] (urging platforms to respect the U.N.'s freedom of expression standards in content moderation); ARTICLE 19, SIDE-STEPPING RIGHTS: REGULATING SPEECH BY CONTRACT 5, 8 (2018), https://bit.ly/2UFY5Xi [https://perma.cc/PW6A-4TQV] (recommending that companies should align their content moderation with international human rights standards and describing how freedom of expression is protected under the ICCPR); Aswad, *supra* note 9, at 67–70 (arguing that the benefits of aligning corporate speech codes with ICCPR standards outweigh the potential downsides).

embodied in U.N. instruments.[11] The UNGPs define corporate "respect" for human rights as meaning that business entities should not only "avoid infringing on the human rights of others" but also "address adverse human rights impacts with which they are involved."[12] As many transnational social media companies are U.S.-based, it should be noted that the U.S. government has called on American companies to treat the UNGPs as a minimum standard in their operations.[13]

Others have raised concerns about the use of the ICCPR's standards in content moderation. For example, some have observed that too much speech would be protected.[14] Several commentators have expressed disquietude that international standards will not provide sufficient guidance to be useful to companies.[15] Others have argued that American companies should use First Amend-

11. Human Rights Council Res. 17/4, U.N. Doc. A/HRC/RES/17/4, ¶ 1 (July 6, 2011); John Ruggie (Special Representative of the Secretary General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises), *Rep. of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises: Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, U.N. Doc. A/HRC/17/31 (Mar. 21, 2011) [hereinafter UNGPs]. The UNGPs define "internationally recognized human rights" as constituting at a minimum the U.N.'s International Bill of Human Rights (which includes the ICCPR) as well as an International Labor Organization Declaration. *Id.* at Principle 12. The commentary notes that additional U.N. instruments may be referred to as well. *Id.* at commentary to Principle 12

12. *Id.* at Principle 11.

13. U.S. DEP'T OF STATE, RESPONSIBLE BUSINESS CONDUCT: FIRST NATIONAL ACTION PLAN FOR THE UNITED STATES OF AMERICA 17 (Dec. 16, 2016), https://bit.ly/3e2cWlV [https://perma.cc/GJ38-QWHJ] ("The U.S. government encourages businesses to treat tools like . . . the U.N. Guiding Principles as a floor rather than a ceiling for implementing responsible business practices . . . .").

14. *See, e.g.*, Brenda Dvoskin, *International Human Rights Law Is Not Enough to Fix Content Moderation's Legitimacy Crisis*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y AT HARV. UNIV. (Sept. 16, 2020), https://bit.ly/3hW3uSj [https://perma.cc/35UD-Q9FF] (noting that international human rights law would protect offensive speech that users would not want to "navigate" on platforms).

15. *See, e.g.*, Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1063 (2018) (dismissing international human rights law as a source of guidance for tech companies in defining hate speech and terrorist-related speech because "human rights law contains exceptionally flexible standards"); Evelyn Douek, *U.N. Special Rapporteur's Latest Report on Online Content Regulation Calls for 'Human Rights by Default*,' LAWFARE (June 6, 2018, 8:00 AM), https://bit.ly/3ATRwkS [https://perma.cc/6U4X-YVVZ] (expressing concern about calls to align content moderation with international human rights law because international human rights law is not a "single, self-contained and cohesive body of rules . . . . [T]hese laws are found in a variety of international and regional treaties that are subject to differing interpretations by states that are parties to the convention as well as international tribunals applying the laws").

ment standards,[16] despite the fact that corporate speech codes have already departed from those standards.[17]

To unpack and analyze the prevailing view that a wide and unbridgeable gap separates the First Amendment's and the ICCPR's approaches to protecting speech, this Article assesses salient interpretations of permissible speech limitations under both bodies of law. Part I engages in an overview of key components of First Amendment law. Part II provides a general overview of U.N. human rights standards for freedom of expression.[18] Parts I and II each contain a section that focuses on hate speech, which is an area of great controversy with respect to content moderation[19] as well as

16. *See, e.g.*, David French, *A Better Way to Ban Alex Jones*, N.Y. TIMES: OP. (Aug. 7, 2018), https://nyti.ms/3yKH11f [https://perma.cc/RA7J-FHBH] (criticizing companies for departing from First Amendment approaches to content curation); *see also* Noah Feldman, *Free Speech Isn't Facebook's Job*, BLOOMBERG: BLOOMBERG OP. (June 1, 2016, 12:08 PM), https://bloom.bg/3yHpbfu [https://perma.cc/SBH3-4F94] (criticizing U.S. social media companies for abandoning First Amendment principles in addressing hate speech on their platforms, but assessing that society cannot expect companies to respect free speech).

17. *See* Ammori, *supra* note 6, at 2274–76 (discussing how various platforms' speech codes are not aligned with First Amendment principles when they forbid hate speech, bullying, sexually explicit content, anonymous speech and other protected expression).

18. Given pervasive misunderstandings about the scope of the U.N.'s global standards on speech and the evolution in the last decade of relevant interpretations by U.N. expert bodies, this Article provides a more intensive examination of U.N. standards than First Amendment jurisprudence.

19. *See, e.g.*, Gilad Edelman, *The Parler Bans Open a New Front in the 'Free Speech' Wars*, WIRED (Jan. 13, 2021, 7:00 AM), https://bit.ly/3yLKSek [https://perma.cc/L7RC-8CS8] (reporting that Silicon Valley titans' ban on social media network Parler for failing to moderate hateful and dangerous speech "opens a new front in the online speech wars"); Dipayan Ghosh, *Op-ed: Why Social Media CEOs Like Zuckerberg and Dorsey Secretly Like Fueling Fake News Debate*, CNBC (May 31, 2020, 9:30 AM), https://cnb.cx/2U1gBJL [https://perma.cc/EL7L-GVDM] (observing that "[t]he social media debate involving free speech, hate speech and disinformation will take years to resolve . . . . The social media content debate has flared up to a level of intensity in recent days that we could never have imagined."); Benjamin Goggin, *YouTube's Week from Hell: How the Debate Over Free Speech Online Exploded After a Conservative Star with Millions of Subscribers Was Accused of Homophobic Harassment*, BUS. INSIDER (June 9, 2019, 1:31 PM), https://bit.ly/3wBZVpw [https://perma.cc/T5PH-MT2H] (noting that "[t]he dispute between Maza and Crowder is just the latest in the emergent tug-of-war between conservative free-speech advocates who often find themselves being accused of hate and harassment and progressive advocates for proactive censorship of what they believe to be hate speech or targeted abuse"); Chris Bell, *The Reddit Boss and the Hate Speech Row*, BBC NEWS (July 10, 2018), https://bbc.in/36tXY3O [https://perma.cc/7QCJ-KCX6] (stating that Reddit's hate speech policy "has reignited a free speech debate among the site's users" and "attracted tens of thousands of reactions and significant discussion").

a topic that is often highlighted to underscore differences between the U.S. and U.N. approaches to expression.[20]

Part III reflects on the analysis of U.S. and U.N. standards and concludes that both bodies of law are grounded in four foundational tenets that discipline the regulation of speech: (1) the presumption in favor of speech, (2) the prohibition of improperly vague or overbroad speech prohibitions, (3) the requirement that restrictions on speech may only be imposed for important public interest objectives, and (4) the necessity of narrowly tailoring burdens on speech. Part III concludes that, while the United States may be an "outlier" on speech compared to approaches taken in the domestic legal systems of other countries, the U.S. approach to freedom of expression is not such an outlier when compared to U.N. protections for freedom of expression, which are more protective of speech than is generally understood to be the case. This Part concludes by reflecting on the implications of this analysis for corporate content moderation.

I. THE U.S. APPROACH TO FREEDOM OF EXPRESSION

Part I(A) provides background on the history and evolution of the First Amendment's protection for freedom of expression. Part I(B) then examines key doctrines that undergird the First Amendment's speech-protective jurisprudence. Part I(C) concludes with a focus on the intersection of this jurisprudence with hate speech.

A. Background

1. History and Evolution

The U.S. approach to freedom of expression is grounded in its Constitution's First Amendment, which went into effect in 1791 and provides that "Congress shall make no law . . . abridging the freedom of speech, or of the press."[21] Constitutional law scholar Erwin Chemerinsky has noted that the "First Amendment undoubtedly was a reaction against the suppression of speech and of the press

20. *See supra* note 2 for scholarly works invoking hate speech rules as driving significant differences between First Amendment jurisprudence and U.N. protections for free speech. Though there is no generally accepted definition of hate speech under international or U.S. law, this Article typically refers to hate speech as "speech that expresses hateful or discriminatory views about certain groups . . . or about certain personal characteristics that have been the basis of discrimination (such as race, religion, gender, and sexual orientation)." STROSSEN, *supra* note 3, at xxiii.

21. U.S. CONST. amend. I.

that existed in English society."[22] He is referring to England's licensing requirements as well as its prohibition on seditious libel.[23] Under the English licensing system, "no publication was allowed without a government-granted license."[24] The seditious libel law criminalized criticism of the government.[25] Chemerinsky has also observed that—beyond ending prior restraints on speech and seditious libel—"there is little indication of what the framers intended."[26] Another scholar has similarly noted that there is "no clear, consistent vision of what the framers meant by freedom of speech."[27]

Given the unclear intent underlying the scope of free speech, it may not be surprising that interpretations of the First Amendment have evolved significantly over time. Despite outrage about the British ban on criticizing the government, the United States swiftly adopted a law in 1798 that criminalized criticism of the government,[28] but it was not challenged in court as a First Amendment violation.[29] When the U.S. Supreme Court decided free speech cases in the 1800s, the Court's "approach was to allow repression of any speech that had a 'bad tendency.'"[30] In the early 1900s, under

22. ERWIN CHEMERINKSY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 1002 (6th ed. 2019).

23. *Id.* at 1002–03.

24. *Id.* at 1002. England's licensing system was in effect until 1694. *Id.*

25. *Id.* at 1002–03. With respect to seditious libel, Chemerinsky notes that "there were fewer prosecutions for seditious libel [in the colonies] than in England, but there were other controls . . . over dissident speech." *Id.* at 1003.

26. *Id.* at 1004.

27. RODNEY A. SMOLLA, SMOLLA AND NIMMER ON FREEDOM OF SPEECH 1–18 (1994); *see also* ANTHONY LEWIS, FREEDOM FOR THE THOUGHT WE HATE: A BIOGRAPHY OF THE FIRST AMENDMENT 10 (2007) ("The birth of the First Amendment threw no light on how its scope should be understood.").

28. LEWIS, *supra* note 27, at 11. The Sedition Act was designed and deployed to prosecute those critical of the President and bolster his chances of gaining reelection. *Id.* at 11–12. Instead, the law helped his Vice President (Thomas Jefferson) win the presidency because the law was invoked to argue that the President sought to return the country to a monarchy. *Id.* at 15.

29. *Id.* Though the statute was not formally challenged in court on First Amendment grounds, many at the time believed it violated free speech protections. *Id.* at 17–18; *see also* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 276 (1964) ("Although the Sedition Act was never tested in this Court, the attack upon its validity has carried the day in the court of history.").

30. LEWIS, *supra* note 27, at 24. Under this test, speech could be banned if it was "contrary to the public welfare." *Id.* If one were to translate this approach through the lens of today's constitutional law rubric, a speech restriction was presumed constitutional, and a challenger would bear the heavy burden of proving that it failed rationale basis review, i.e., that it had no legitimate goal and/or was irrational in terms of promoting the goal. Such an approach directly contrasts with the current approach in which all speech restrictions are subject to heightened scrutiny. *See infra* Part I.B. The Supreme Court essentially viewed the point of the

this approach, the Supreme Court upheld the prosecution of Americans during World War I for protesting the war effort.[31] It was not until 1931 that the Supreme Court overturned a law on free speech grounds.[32] After several more decades, the Supreme Court developed a highly speech-protective jurisprudence, which this Article describes in Part I(B). But the road to providing some of the broadest speech protections in history was a long one, paved with decades of lawful suppression of speech—including political views and human rights activism—solely because of speculative fears that such speech might indirectly lead to some potential future harm.

Despite scholarly assessments that the framers' intent was unclear and a history of punishing controversial speech until the 1960s, the United States has often projected to the world a myth that the First Amendment has from its enactment provided broad speech protections. For example, when defending the U.S. decision not to ban a high profile video that offended religious sensibilities, President Obama explained to the U.N. General Assembly that "our founders understood that without [broad] protections [for hateful and offensive speech], the capacity of each individual to express their own views . . . may be threatened."[33] At the same time, in responding to international criticism that it is not amenable to exploring or engaging with other standards for speech, the United States has invoked how its unfortunate experience in suppressing speech has profoundly shaped its views on expression, including why it believes non-censorial, good governance measures are more appropriate (and effective) than speech bans to solve a range of issues.[34]

First Amendment as outlawing prior restraints on speech rather than addressing punishments for speech once expressed. LEWIS, *supra* note 27, at 24.

31. *Id.* at 25–28. Americans were prosecuted under the Espionage Act for distributing leaflets that compared conscription to slavery, discussing socialism and interactions with inmates who were convicted for failing to register for the draft, and distributing leaflets that criticized the President's decision to send troops into Russia. *Id.*

32. *Id.* at 39.

33. Barack Obama, President, United States, Remarks by the President to the U.N. General Assembly (Sept. 25, 2012, 10:22 AM), https://bit.ly/3qZS8k4 [https://perma.cc/RUU7-ZXXM].

34. *See, e.g.*, PERMANENT MISSION OF THE U.S., UNITED STATES GOVERNMENT RESPONSE TO THE UNITED NATIONS OFFICE OF THE HIGH COMMISSIONER FOR HUMAN RIGHTS CONCERNING EXPERT WORKSHOPS ON INCITEMENT TO NATIONAL, RACIAL OR RELIGIOUS HATRED 1 (Nov. 3, 2010), https://bit.ly/3ARV8DS [https://perma.cc/SK2Q-72K3] (explaining the United States' reluctance to ban hateful speech is based in part on the American experience of suppressing speech and endorsing pro-active, non-censorial methods to combat national, racial, or religious intolerance); *see also* U.S. GOV'T, PERIODIC REPORT OF THE UNITED STATES OF AMERICA TO THE UNITED NATIONS COMMITTEE ON THE ELIMINATION

2. *Overview*

Several key principles form the foundational framework that animates the First Amendment's contemporary approach to freedom of speech.[35] First, any governmental restriction of speech is void if it is either unduly vague or overly broad.[36] Second, the First Amendment requires that speech infringements be narrowly tailored to achieve a governmental interest that is at least substantial, which cannot include favoring or disfavoring particular messages.[37] If a speech restriction is "content-based," then a court will apply "strict scrutiny" in assessing the law, which means a court must find that the law is the least restrictive means to advance a compelling governmental interest.[38] If a speech restriction is content-neutral, then a court will apply "intermediate scrutiny" to assess if the law is appropriately tailored to achieve a substantial governmental interest.[39] Third, the Supreme Court has also made a normative determination that certain categories of speech are unprotected and, therefore, the government may regulate such speech more liberally.[40] However, even with respect to those categories of unprotected expression, the government may not regulate such speech in a content-discriminatory manner.[41]

B. Key Principles

1. *Vagueness and Overbreadth*

One of the fundamental principles of the First Amendment's approach to speech is that any prohibitions of expression may not be unduly vague.[42] A speech ban "is unconstitutionally vague if a reasonable person cannot tell what speech is prohibited and what is

---

OF RACIAL DISCRIMINATION CONCERNING THE INTERNATIONAL CONVENTION ON THE ELIMINATION OF ALL FORMS OF RACIAL DISCRIMINATION 30–31 (2013), https://bit.ly/3xwWMsw [https://perma.cc/E5D6-CKZ4] (explaining the United States' experience in banning speech and why the United States does not have a blanket ban on racial hate speech).

35. *See* CHEMERINSKY, *supra* note 22, at 1012 (observing that key doctrines in First Amendment methodology include the prohibition on vague and over broad laws as well as viewpoint discrimination).

36. *See infra* notes 42–58 and accompanying text.

37. *See infra* notes 59–81 and accompanying text.

38. *See infra* notes 66–68 and accompanying text.

39. *See infra* notes 74–76 and accompanying text.

40. *See infra* notes 82–84 and accompanying text.

41. *See infra* notes 103–07 and accompanying text.

42. *See* NAACP v. Button, 371 U.S. 415, 432–33 (1963) ("[S]tandards of permissible statutory vagueness are strict in the area of free expression . . . . Because First Amendment freedoms need breathing space to survive, government may regulate in the area only with narrow specificity.") (citations omitted).

permitted."[43] This doctrine is based on fairness (i.e., people should have notice of what words or conduct violate the law), the concern that broad laws give government implementers too much discretion and risk discriminatory prosecutions, and the fear of chilling protected expression.[44] Indeed, Chemerinsky calls the bar on unduly vague speech prohibitions "a *powerful* tool in First Amendment litigation because it allows facial challenges to laws even by those whose speech otherwise would be unprotected by the First Amendment."[45]

The U.S. Supreme Court has struck down a variety of speech bans as unduly vague. For example, the Court invalidated a Massachusetts statute that criminalized publicly treating the U.S. flag "contemptuously" for "fail[ing] to draw reasonably clear lines between the kinds of nonceremonial treatment that are criminal and those that are not."[46] Similarly, the Court held as unduly vague a Washington state law that required government employees take an oath swearing they were not "subversive" persons because it failed to give notice of what expressions were outlawed.[47] The Court has also held as unconstitutionally vague a California law prohibiting the display of certain signs "of opposition to organized government"[48] and a local ordinance in Ohio that criminalized the assembly of persons who "conduct themselves in a manner annoying to persons passing by."[49] Lower courts have similarly invalidated speech restrictions on vagueness grounds.[50]

---

43. CHEMERINSKY, *supra* note 22, at 1025 (citing to Connally v. Gen. Constr. Co., 269 U.S. 385, 391 (1926), which stated a law is unduly vague if "[people] of common intelligence must necessarily guess at its meaning").

44. *Id.* at 1025–26.

45. *Id.* at 1027 (emphasis added); *see also* DAVID A. FARBER, THE FIRST AMENDMENT 55 (5th ed. 2019) ("A badly drafted regulation can be struck down on its face without any inquiry into its application to the particular plaintiff challenging the regulation.").

46. Smith v. Goguen, 415 U.S. 566, 574 (1974). The Court further noted that "[s]tatutory language of such a standardless sweep allows policemen, prosecutors, and juries to pursue their personal predilections. Legislatures may not so abdicate their responsibilities for setting the standards of the criminal law." *Id.* at 575.

47. *See* Baggett v. Bullitt, 377 U.S. 360, 372 (1964) (finding that the oath does not "provide[ ] an ascertainable standard of conduct").

48. Stromberg v. California, 283 U.S. 359, 361, 369–70 (1931).

49. Coates v. City of Cincinnati, 402 U.S. 611, 611, 614 (1971).

50. *See, e.g.*, Ketchens v. Reiner, 239 Cal. Rptr. 549, 553–54 (Cal. Ct. App. 1987) (finding that a plaintiff seeking a preliminary injunction was likely to succeed in arguing that a criminal ban on insulting and abusing teachers was impermissibly vague); Gatto v. Cnty. of Sonoma, 120 Cal. Rptr. 2d 550, 573–74 (Cal. Ct. App. 2002) (finding a prohibition on clothing "intended to provoke, offend, or intimidate others . . . including offensive slogans, insignia or 'gang colors'" was unduly vague).

Related to the prohibition of vague speech bans is the bar on overly broad speech restrictions, which is triggered when laws "regulate substantially more speech than the Constitution allows to be regulated."[51] The Supreme Court has struck down a variety of speech bans based on overbreadth. For example, the Court overturned as overbroad a ban on "opprobrious words or abusive language, tending to cause a breach of the peace,"[52] a local ordinance in New Jersey that banned not only nude dancing but also all live entertainment,[53] a local law in Texas that made it unlawful to "interrupt any policeman in the execution of his duty,"[54] and a rule prohibiting all "First Amendment activities" within a certain area of the Los Angeles Airport.[55] More recently, the Court has applied this doctrine to invalidate a ban on depictions of animal cruelty[56] as well as a North Carolina law prohibiting convicted sex offenders from accessing social media.[57] Lower courts have also applied this principle to invalidate speech bans.[58]

2. *Content-Neutrality and Narrow Tailoring*

A cardinal principle in First Amendment jurisprudence is that the "government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[59] The initial step in implementing this content-neutrality principle is to determine if a speech restriction is both "viewpoint neutral" and "subject matter neutral."[60] A law is not *viewpoint neutral* if it regulates

51. CHEMERINSKY, *supra* note 22, at 1027. In other words, overbreadth is distinguishable from vagueness because a law can be precise in its phrasing but still be overly broad in prohibiting otherwise protected speech.

52. Gooding v. Wilson, 405 U.S. 518, 519, 527–28 (1972).

53. Schad v. Borough of Mount Ephraim, 452 U.S. 61, 76 (1981).

54. City of Houston v. Hill, 482 U.S. 451, 455, 467 (1987). The Court noted such a law is violated every day, but "only some individuals—those chosen by the police in their unguided discretion—are arrested." *Id.* at 466–67. Such selective prosecution "is susceptible of regular application to protected expression." *Id.* at 467.

55. Bd. of Airport Comm'rs of L.A. v. Jews for Jesus, Inc., 482 U.S. 569, 570, 577 (1987).

56. United States v. Stevens, 559 U.S. 460, 482 (2010).

57. Packingham v. North Carolina, 137 S. Ct. 1730, 1738 (2017).

58. *See, e.g.*, Van Nuys Publ'g Co. v. City of Thousand Oaks, 489 P.2d 809, 810 (Cal. 1971) (holding that a ban on placing literature on someone else's property without consent was overly broad); Welton v. City of Los Angeles, 556 P.2d 1119, 1124–25 (Cal. 1976) (deciding that a city ordinance banning sidewalk sales of "merchandise" to be overbroad); Prigmore v. City of Redding, 150 Cal. Rptr. 3d 647, 665 (Cal. Ct. App. 2012) (finding a local policy that "bans *all* leafleting involving the solicitation of funds" in libraries to be overly broad).

59. Police Dep't of Chi. v. Mosley, 408 U.S. 92, 95 (1972).

60. CHEMERINSKY, *supra* note 22, at 1014–15.

speech based on the ideas or philosophies expressed.[61] For example, the Supreme Court has held that a federal statute that prohibited trademarks that could "disparage . . . or bring . . . into contemp[t] or disrepute" any "persons, living or dead" to constitute viewpoint discrimination.[62]

A law is not *subject matter neutral* if it regulates "speech based on the topic of the speech."[63] The Court has provided a framework for determining whether a law is content-based, which involves reviewing the text of a speech restriction as well as governmental motives for its enactment.[64] Under the Court's precedents, content-based laws have included those which regulated all speech but excluded labor-related speech as well as those which regulated only sexual speech.[65]

Under the First Amendment, "content-based restrictions on speech [are] presumed invalid" and "the Government bear[s] the burden of showing their constitutionality."[66] In other words, al-

61. *Id.* at 1014 ("Viewpoint neutral means that the government cannot regulate speech based on the ideology of the message.").

62. Matal v. Tam, 137 S. Ct. 1744, 1751 (2017). In *Matal*, an Asian American rock band sought a trademark for its name: The Slants. *Id.* at 1754. While "Slants" is a derogatory reference to Asian Americans, the band members believed that by embracing the term, they could remove its stigma and "reclaim" it. *Id.* In finding the federal statute to constitute impermissible viewpoint discrimination, the Court reiterated, "We have said time and again that 'the public expression of ideas may not be prohibited merely because the ideas are themselves offensive to some of their hearers.'" *Id.* at 1763 (citing Street v. New York, 394 U.S. 576, 592 (1969)).

63. CHEMERINSKY, *supra* note 22, at 1015.

64. *See* Reed v. Town of Gilbert, 576 U.S. 155, 163–64 (2015). This framework encompasses the following three areas of inquiry. First, if a law is content-based on its face, then courts should presume it is unconstitutional and the government must overcome strict scrutiny to justify its actions. *Id.* Second, if the law is facially content-neutral but cannot be "justified without reference to the content of the regulated speech," then it will also be subject to strict scrutiny. *Id.* at 164 (quoting Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989)). Third, if the law is content-neutral on its face, but it was "adopted because of the disagreement with the message [the expression] conveys," then the court will use strict scrutiny to assess the speech restriction. *Id.* (quoting *Ward*, 491 U.S. at 791). The last two lines of inquiry focus on the motive for the governmental intrusion on speech. It should be noted that in limited cases that pre-dated *Reed*, the Court allowed governmental authorities to counter a finding that a regulation is content-based "by persuading a court that the regulation is justified by a content-neutral desire to avoid undesirable secondary effects of speech." CHEMERINSKY, *supra* note 22, at 1020.

65. *See* Carey v. Brown, 447 U.S. 455, 470–71 (1980) (finding that an Illinois statute which banned peaceful picketing in residential neighborhoods but allowed for labor protests violated content- neutrality for favoring one particular topic: labor grievances); United States v. Playboy Ent. Grp. Inc., 529 U.S. 803, 811–12, 826–27 (2000) (holding that a federal statute focused on "sexually explicit adult programming" and channels "primarily dedicated to sexually-oriented programming" constitutes subject matter regulation inconsistent with content–neutrality).

66. Ashcroft v. ACLU, 542 U.S. 656, 660 (2004).

though content-based restrictions are inherently suspect, they are not automatically void. Instead, the Supreme Court subjects content-based restrictions to "strict scrutiny" to determine their validity.[67] Strict scrutiny means the government bears the burden of demonstrating that the speech restriction (1) is necessary (i.e., constitutes the least restrictive means) to (2) achieve a compelling governmental interest.[68]

The Supreme Court has almost invariably found content-based speech restrictions to fail strict scrutiny. Typically, the Court has held that such restrictions do not reflect the least intrusive means of achieving a compelling governmental interest.[69] The Court has also found that the government failed the strict scrutiny test because the

---

67. *See, e.g.*, Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 658–59 (1994) (noting that strict scrutiny is applied to speech restrictions that "reflect the Government's preference for the substance of what the favored speakers have to say (or aversion to what the disfavored speakers have to say)").

68. *See Playboy Ent. Grp.*, 529 U.S. at 813 ("If a statute regulates speech based on its content, it must be narrowly tailored to promote a compelling Government interest. If a less restrictive alternative would serve the Government's purpose, the legislature must use that alternative." (citation omitted)); Sable Commc'ns of Cal., Inc. v. FCC, 492 U.S. 115, 126 (1989) ("The Government may . . . regulate the content of constitutionally protected speech in order to promote a compelling interest if it chooses the least restrictive means to further the articulated interest."). The Court has elaborated on the least restrictive means test by noting:

> The purpose of the test is not to consider whether the challenged restriction has some effect in achieving Congress' goal, regardless of the restriction it imposes. The purpose of the test is to ensure that speech is restricted no further than necessary to achieve the goal, for it is important to ensure that legitimate speech is not chilled or punished. For that reason, the test does not begin with the status quo of existing regulations, then ask whether the challenged restriction has some additional ability to achieve Congress' legitimate interest. Any restriction on speech could be justified under that analysis. Instead, the court should ask whether the challenged regulation is the least restrictive means among available, effective alternatives.

Ashcroft, 542 U.S. at 666.

69. *See, e.g.*, *Ashcroft*, 542 U.S. at 659–60, 666–67 (finding a federal statute that required sexually oriented websites to conduct age verification constituted a content-based restriction and likely would fail the least restrictive alternative test because users could install filters on their own devices to protect children from sexual material); *Playboy Ent. Grp*., 529 U.S. at 823 (holding the government failed to demonstrate that daytime ban on expression was the least restrictive means when plausible less intrusive alternatives existed and noting "[i]t was for the Government, presented with a plausible, less restrictive alternative, to prove the alternative to be ineffective"); United States v. Alvarez, 567 U.S. 709, 728–29 (2012) (plurality opinion) (finding that a federal criminal ban on lying about receiving military honors was not the least restrictive means to protect the integrity of the awards system because, among other things, counter-speech and the creation of an Internet database of awardees would achieve the objective without burdening speech).

justification for burdening speech was not directly linked to the harm or injury it intended to prevent.[70] At times, the Court has rejected as pretextual the government's assertion of a compelling public purpose.[71] The Court has also found that certain governmental interests do not reach the level of being "compelling."[72] None-

70. *See, e.g.*, *Alvarez,* 567 U.S. at 725 (finding that "There must be a direct causal link between the restriction imposed and the injury to be prevented" and the government failed to show that its interest in "protecting the integrity of the military honors system" was directly linked to criminalizing "false claims of liars" about receiving such military awards); Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 799 (2011) (finding a California law invalid because it restricted the sale of violent video games to minors when there was no direct causal link between actual harm to minors and the violent video games); *see also* Clay Calvert & Matthew D. Bunker, *An Actual Problem in First Amendment Jurisprudence: Examining the Immediate Impact of Brown's Proof-of-Causation Doctrine on Free Speech and Its Compatibility with the Marketplace Theory*, 35 HASTINGS COMMC'NS. & ENT. L.J. 391, 404 (2013) ("[T]he phrase 'direct causal link' is brand new within the Supreme Court's First Amendment jurisprudence, having only entered the doctrinal lexicon in *Brown* and *Alvarez*.").

71. *See, e.g.*, Williams Yulee v. Fla. Bar, 575 U.S. 433, 448 (2015) (plurality opinion) (citing *Brown*, 564 U.S. at 802). The Court noted "that underinclusiveness [of a speech restriction] can raise 'doubts about whether the government is in fact pursuing the interest it invokes, rather than disfavoring a particular speaker or viewpoint.'" *Id.* It explained that "[i]n a textbook illustration of that principle, we invalidated a city's ban on ritual animal sacrifices because the city failed to regulate vast swaths of conduct that similarly diminished its asserted interests in public health and animal welfare." *Id.* (citing Church of Lukumi Babalu Aye, Inc. v. Hialeah, 508 U.S. 520, 543–47 (1993)). The Court also highlighted that "underinclusiveness can also reveal that a law does not actually advance a compelling interest. For example, a State's decision to prohibit newspapers, but not electronic media, from releasing the names of juvenile defendants suggested that the law did not advance its stated purpose of protecting youth privacy." *Id.* at 449 (citing Smith v. Daily Mail Publ'g Co., 443 U.S. 97, 104–05 (1979)).

72. *See* Cohen v. California, 403 U.S. 15, 26 (1971) (noting "absent a *more particularized* and compelling reason for its actions, the State may not, consistently with the First and Fourteenth Amendments, make the simple public display here involved of this single four-letter expletive a criminal offense." (emphasis added)); *see also* Members of City Council of Los Angeles v. Taxpayers for Vincent, 466 U.S. 789, 823 (1984) (Brennan, J., dissenting) (explaining that "a governmental interest in aesthetics cannot be regarded as *sufficiently* compelling to justify a restriction of speech based on an assertion that the content of the speech is, in itself, aesthetically displeasing." (emphasis added)). Additionally,

> [T]he U.S. Supreme Court has said that '[w]here the designed benefit of a content-based speech restriction is to shield the sensibilities of listeners, the general rule is that the right of expression prevails, even where no less restrictive alternative exists,' implicitly holding that there is no compelling state interest based solely on protecting those who are 'sensitive.'

Sarah E. Smith, Comment, *Threading the First Amendment Needle: Anonymous Speech, Online Harassment, and Washington's Cyberstalking Statute*, 93 WASH. L. REV. 1563, 1573 (2018) (quoting *Playboy Ent. Grp.*, 529 U.S. at 813)).

theless, a few content-based restrictions have survived the strict scrutiny test.[73]

If a speech restriction qualifies as content-neutral (i.e., is not content-based),[74] then courts apply an intermediate scrutiny test.[75] To survive this level of scrutiny, the government must demonstrate (1) the restriction on speech does "not 'burden substantially more speech than is necessary to further'" (2) an important governmental interest.[76] The Court has emphasized that the government "may not regulate expression in such a manner that a substantial portion of the burden on speech does not serve to advance its goals."[77] Though commentators have noted that the line between "compelling" and other important governmental purposes remains unclear,[78] the Court's requirement for narrowly tailoring and

73. *See, e.g.*, *Williams Yulee*, 575 U.S. at 455 (finding that Florida law outlawing judicial candidates from "personally soliciting campaign funds" survived strict scrutiny's least intrusive means test for achieving the goal of preserving public perception of the integrity of the state judiciary); Burson v. Freeman, 504 U.S. 191, 211 (1992) (plurality opinion) (holding that a Tennessee law that restricts solicitation of votes to 100 feet away from polling place entrances in order to preserve election integrity passes the strict scrutiny test); Holder v. Humanitarian L. Project, 561 U.S. 1, 39 (2010) (finding a material support statute to survive strict scrutiny with respect to particular activities).

74. Turner Broad. Sys. v. FCC, 512 U.S. 622, 643 (1994) (observing that "laws that confer benefits or impose burdens on speech without reference to the ideas or views expressed are in most instances content neutral."); *Taxpayers for Vincent*, 466 U.S. at 804 (finding that a local law forbidding signs on public property was neutral about "any speaker's point of view"); Heffron v. Int'l Soc'y for Krishna Consciousness, Inc., 452 U.S. 640, 648–49 (1981) (finding that a local law designating the locations where solicitations can occur was applicable fairly to all regardless of viewpoint). The Court has also noted that "we have often declared that '[a] state or municipality may protect individual privacy by enacting reasonable time, place, and manner regulations applicable to all speech *irrespective of content*.'" Carey v. Brown, 447 U.S. 455, 470 (1980) (citing Erznoznik v. City of Jacksonville, 422 U.S. 205, 209 (1975)).

75. *Turner Broad. Sys.*, 512 U.S. at 642 ("[R]egulations that are unrelated to the content of speech are subject to an intermediate level of scrutiny because in most cases they pose a less substantial risk of excising certain ideas or viewpoints from the public dialogue." (citation omitted)).

76. McCullen v. Coakley, 573 U.S. 464, 486 (2014) (quoting Ward v. Rock Against Racism, 491 U.S. 781, 799 (1989)). The significance of the governmental objective has been defined by the Supreme Court as "important" or "substantial." *Turner Broad. Sys.*, 512 U.S. at 662 (quoting United States v. O'Brien, 391 U.S. 367, 377 (1968)).

77. *McCullen*, 573 U.S. at 486 (quoting *Ward*, 491 U.S. at 799).

78. *See, e.g.*, Richard H. Fallon, Jr., *Strict Judicial Scrutiny*, 54 UCLA L. REV. 1267, 1321 (2007) (observing that "the Supreme Court has frequently adopted an astonishingly casual approach to identifying compelling interests"); Stephen E. Gottlieb, *Compelling Governmental Interests: An Essential but Unanalyzed Term in Constitutional Adjudication*, 68 B.U. L. REV. 917, 941 (1988) (assessing that "the governmental interests identified by both the full Court and its individual members . . . indicates that . . . their sources are ambiguous and their relative weights impos-

"demanding a *close fit* between ends and means . . . prevents the government from too readily 'sacrific[ing] speech for efficiency.'"[79] A number of content-neutral speech restrictions have survived intermediate scrutiny,[80] but some have not.[81]

3. Categories of Unprotected Speech

The Supreme Court has identified certain "historic and traditional categories" of speech as unprotected, including obscenity,

---

sible to gauge"); *Let the End Be Legitimate: Questioning the Value of Heightened Scrutiny's Compelling-and Important-Interest Inquiries*, 129 HARV. L. REV. 1406, 1409 (2016) (arguing that the Court "rarely deals in depth with the state-interest question" because it often disposes of a case based on narrow tailoring issues and when "the Court *has* ruled on the state-interest question . . . such rulings bear little on . . . the challenged state action's constitutionality").

79. *McCullen*, 573 U.S. at 486 (emphasis added) (quoting Riley v. Nat'l Fed'n of the Blind of N.C., Inc., 487 U.S. 781, 795 (1988)).

80. *See, e.g.*, Frisby v. Schultz, 487 U.S. 474, 488 (1988) (finding a local ordinance which prohibited picketing in front of residences or dwellings to be constitutional); City of Renton v. Playtime Theaters, Inc., 475 U.S. 41, 48, 54–55 (1986) (holding a zoning ordinance that restricted location of adult theatres to be content-neutral because it was dealing with the secondary effects of the theatres and that it survived constitutional scrutiny).

81. *See, e.g.*, *McCullen*, 573 U.S. at 485, 497 (finding that a non-content-based Massachusetts statute that required a buffer zone near locations where abortions are performed was not narrowly tailored and failed constitutional scrutiny); Packingham v. North Carolina, 137 S. Ct. 1730, 1736–37 (2017) (finding that a statute prohibiting sex offenders from accessing social media was not narrowly tailored to achieve the government's objective, even if the statute was assumed to be content-neutral); McCraw v. City of Oklahoma City, 973 F.3d 1057, 1070–80 (10th Cir. 2020) (finding a city ordinance that prohibited staying on medians was not narrowly tailored and the government could deploy less burdensome means to achieve safety goals). Before departing from the topic of strict and intermediate scrutiny, it should be noted that the Supreme Court has "treated speech in some government places differently based on the need for greater governmental control." CHEMERINSKY, *supra* note 22, at 1229. For example, "public forums" are publicly owned areas—such as parks—that the government is required "to make available for speech." *Id.* at 1233. Within a public forum, content-based regulations remain subject to strict scrutiny and content-neutral regulations continue to be subject to intermediate scrutiny. *Id.* (This analysis also applies to government-owned property that, although not traditionally viewed as a public forum, is designated as such by the government. *Id.* at 1244.) Within government-created limited or non-public forums, speech may be restricted in ways that are not permissible in public forums. *Id.* at 1245–46. For example, the Supreme Court has noted that, with respect to a limited public forum, the government is not required to allow all types of speech but must not engage in viewpoint discrimination, and its restriction should be reasonable when compared to the forum's purpose. Good News Club v. Milford Cent. Sch., 533 U.S. 98, 106–07 (2001). This Article focuses on comparing First Amendment and U.N. free speech approaches to laws of general applicability in order to assess their viability in the context of content moderation of transnational platforms that are not limited in scope to particular purposes or topics and cover vast numbers of users. The Article therefore does not delve into the particularities of applicable rules for limited public forums.

fraud, defamation, incitement, and speech integral to criminal conduct.[82] The Court has dismissed governmental arguments to expand this listing based on a balancing of the "value of the speech against its societal costs," noting such a "free-floating test" would be "dangerous."[83] Rather, the Court has limited this category of unprotected content-based speech to historically unprotected speech.[84] Though there are several categories of speech that are unprotected because they directly cause specific harm, this Section focuses on those that are most relevant to the hate speech discussion in Part I(C).

The first category is advocacy of incitement to illegal action, such as violence. In 1969, the Supreme Court overturned a conviction of a Ku Klux Klan leader for advocating unlawful means to achieve his group's goals based on, *inter alia*, racist speech, a display of firearms, and a statement that it was "possible that there might have to be some revengeance taken" if white suppression continues.[85] The Court determined that an incitement conviction meets constitutional standards if the government can prove (1) the speech is likely to cause (2) imminent lawless or violent action and (3) the speaker intended to cause the imminent lawless or violent action, which was not the case in *Brandenburg*.[86] Using this standard, the Court has dismissed convictions based on language as well as context that did not constitute a call for imminent lawlessness[87] and has made clear that "mere *advocacy* of the use of force or violence does not remove speech from the protection of the First Amendment."[88]

Related to the issue of speech that calls for violence is the category of "true threats," which the Supreme Court has held is a type of speech that the government may prohibit.[89] True threats "encompass those statements where the speaker means to communicate a

82. United States v. Stevens, 559 U.S. 460, 468–69 (2010).

83. *Id.* at 470.

84. *See* Nadine Strossen, *United States v. Stevens: Restricting Two Major Rationales of Content-Based Restrictions*, 2010 CATO SUP. CT. REV. 67, 82–83 (noting the Court's strict approach to a historical grounding for unprotected speech and that such an approach would preclude hate speech from falling within unprotected categories of speech).

85. Brandenburg v. Ohio, 395 U.S. 444, 445–46 (1969) (per curiam).

86. *Id.* at 447–49.

87. Hess v. Indiana, 414 U.S. 105, 107–09 (1973) (per curiam) (overturning a disorderly conduct conviction for saying either "'[w]e'll take the fucking street later,' or '[w]e'll take the fucking street again'" in the context of police clearing the streets of protestors at an anti-war rally).

88. NAACP v. Claiborne Hardware Co., 458 U.S. 886, 927 (1982).

89. Virginia v. Black, 538 U.S. 343, 359 (2003) (plurality opinion) (noting "the First Amendment . . . permits a State to ban a 'true threat'" (quoting Watts v. United States, 394 U.S. 705, 708 (1969))).

serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals."[90] The point of allowing bans on true threats is to shield "'individuals from the fear of violence' and 'from the disruption that fear engenders,' in addition to protecting people 'from the possibility that the threatened violence will occur.'"[91] Not all intimidating speech will qualify as a true threat: only speech "where a speaker directs a threat to a person or group of persons with the intent of placing the victim in fear of bodily harm or death" will qualify as a true threat.[92]

The Supreme Court determined in 1942 that "fighting words" are also outside of the First Amendment's protection.[93] It defined fighting words as "those which by their very utterance inflict injury or tend to incite an immediate breach of the peace."[94] Although the Court has not formally overturned this decision, it has reversed every fighting words conviction that has come before it since 1942.[95] The Court has found that fighting words laws have been unconstitutional, *inter alia*, because they cannot survive vagueness or overbreadth challenges as well as scrutiny under rules relating to content-based restrictions.[96]

---

90. *Black*, 538 U.S. at 359. In this case, the Court held that Virginia's prohibition on cross burning with intent to intimidate was constitutional but making cross burning *prima facie* evidence of an intent to intimidate was not. *Id.* at 363–64.

91. *Id.* at 360 (quoting R.A.V. v. City of St. Paul, 505 U.S. 377, 388 (1992)).

92. *Id.* Chemerinsky notes that there is a circuit split about whether the perspective of the "reasonable listener" or the "reasonable speaker" should be used to assess if a true threat has been made. CHEMERINSKY, *supra* note 22, at 1090.

93. Chaplinsky v. New Hampshire, 315 U.S. 568, 573–74 (1942). In *Chaplinsky*, a member of the Jehovah's Witnesses had violated a New Hampshire law stating:

> No person shall address any offensive, derisive or annoying word to any other person who is lawfully in any street or other public place, nor call him by any offensive or derisive name, nor make any noise or exclamation in his presence and hearing with intent to deride, offend or annoy him, or to prevent him from pursuing his lawful business or occupation.

Id. at 569. The speech that violated this provision was as follows: "'You are a God damned racketeer' and 'a damned Fascist and the whole government of Rochester are Fascists or agents of Fascists.'" *Id.* The Court found that "face-to-face" statements like "'damn racketeer' and 'damn Fascist' are epithets likely to provoke the average person to retaliation, and thereby cause a breach of the peace" and upheld the conviction. *Id.* at 573–74.

94. *Id.* at 572.

95. CHEMERINSKY, *supra* note 22, at 1094.

96. *Id.* Chemerinsky notes that the Supreme Court has also "narrowed the scope of the fighting words doctrine by ruling that it applies only to speech directed at another person that is likely to produce a violent response." *Id.* Thus, offensive language aimed at an entire group rather than a particular person would not qualify as "fighting words." *But see* David L. Hudson, *The Fighting Words Doctrine: Alive and Well in the Lower Courts*, 19 UNIV. N.H. L. REV. 1, 6–17

C. Hate Speech

Though hate speech bans generally do not survive First Amendment scrutiny, it is interesting to note that in the mid-20th century the Supreme Court had opened the door to such bans. A 1952 Supreme Court case upheld a ban on racist or religious hate speech, which prohibited portraying the "depravity, criminality, unchastity, or lack of virtue of a class of citizens, of any race, color, creed, or religion" that would expose those individuals "to contempt, derision, or obloquy."[97] Though this decision upheld a ban on "group defamation," subsequent cases have rendered it unsustainable as legal authority to support hate speech bans.[98]

As noted by leading First Amendment author and Professor Emerita Nadine Strossen, hate speech bans "are plagued by vagueness and overbreadth, [which] pose virtually unsurmountable challenges."[99] She highlights that U.S. courts have repeatedly found bans on hateful, insulting, and offensive speech to fail based on vagueness and/or overbreadth grounds, including a federal law that prohibited speech that threatened the dignity of foreign embassy workers, noting that a "'dignity' standard, like the 'outrageousness' standard . . . is so inherently subjective that it would be inconsistent with" the Court's approach of not punishing speech that may have an adverse impact on listeners.[100] Professor Strossen also observes that courts have systematically struck down campus hate speech bans as void for vagueness and/or overbreadth.[101] She highlights as an example the University of Michigan's code, which contained terms such as "stigmatize," "victimize," and "interfering with an in-

(2020) (describing examples of the fighting words doctrine surviving scrutiny in lower courts despite Supreme Court jurisprudence that would indicate the demise of this unprotected category of speech).

97. Beauharnais v. Illinois, 343 U.S. 250, 251, 266–67 (1952).

98. CHEMERINSKY, *supra* note 22, at 1104 (noting a variety of reasons for which *Beauharnais* is no longer good law, including that it was based on premises involving defamation that were later overturned and that it violated vagueness/overbreadth principles as well as those on viewpoint discrimination). In addition, *Beauharnais* was premised on the Court's bad tendency rationale for banning speech, which pre-dated its *Brandenburg* decision. See *supra* notes 85–88 and accompanying text for a discussion of *Brandenburg*.

99. STROSSEN, *supra* note 3, at 105.

100. *Id.* at 72 (referring to Boos v. Berry, 485 U.S. 312 (1988)).

101. *Id.* at 77; *see also* CHEMERINSKY, *supra* note 22, at 1106 (noting "the federal courts that thus far have considered the constitutionality of university hate speech codes have invalidated them on vagueness and overbreadth grounds."); Speech First, Inc. v. Fenves, 979 F.3d 319, 338–39 (5th Cir. 2020) (noting that a "consistent line of cases . . . have uniformly found campus speech codes unconstitutionally overbroad or vague").

dividual's academic efforts" that did not survive the vagueness test in court.[102]

The principles of viewpoint and subject matter neutrality also serve to limit potential hate speech bans. In 1992, the Supreme Court overturned a local ordinance that prohibited certain speech "which one knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender."[103] The lower court had narrowed the ordinance to cover only "fighting words," an unprotected category of speech.[104] The Court noted that "these [unprotected] areas of speech can, consistently with the First Amendment, be regulated *because of their constitutionally proscribable content* . . . [but they are not] categories of speech entirely invisible to the Constitution, so that they may be made the vehicles for content discrimination unrelated to their distinctively proscribable content."[105] The Court noted that the law in question prohibited "abusive invective" based on limited grounds such as race, religion, or gender but did not prohibit similar speech on the basis of political opinion, sexual orientation, or union membership and discriminated on the basis of viewpoint.[106] It held this law to be a violation of the viewpoint and subject matter neutrality principles (with respect to a category of

102. STROSSEN, *supra* note 3, at 77. It should also be noted that the Supreme Court has narrowed the space for liability for hateful and offensive speech that may be sought under the tort of intentional infliction of emotional distress. The Supreme Court has held that this tort could not be used to impose liability on protestors near a funeral for highly disturbing speech because, among other things, their speech covered matters of public concern and was delivered at a public place near a public street. Snyder v. Phelps, 562 U.S. 443, 454–455 (2011). The Court also found that the tort's requirement that the picketing be "outrageous" was an overly malleable concept which risked misuse and could not survive First Amendment scrutiny. *Id.* at 458. Prior Supreme Court cases have also narrowed the space for seeking damages for hateful or offensive speech through this tort. *See* Nadine Strossen, *Regulating Racist Speech on Campus: A Modest Proposal?*, 1990 DUKE L.J. 484, 516–517 (explaining that Supreme Court case law imposes high standards for invocations of this tort by public officials and, because the tort relies on a finding of the "outrageousness" of the speech, such vague and broad terminology fails to meet First Amendment scrutiny).

103. R.A.V. v. City of St. Paul, 505 U.S. 377, 380, 391 (1992).

104. *Id.* at 380–81. For a discussion of fighting words, see *supra* notes 93–96. The Supreme Court in *R.A.V.* neither implemented the petitioner's (and amicis') request to revisit the fighting words exception, nor did the Court substantively reaffirm that exception. *Id.* at 381. Instead, it assumed that fighting words are punishable, but nonetheless determined the law to be unconstitutional because of content neutrality principles. *Id.*

105. *Id.* at 383–84. The Court provided the following example: "the government may proscribe libel; but it may not make the further content discrimination of proscribing *only* libel critical of the government." *Id.* at 384.

106. *Id.* at 391.

unprotected speech) because the law was unconstitutionally selective.[107]

The requirement of viewpoint neutrality, even with respect to unprotected speech, provides an additional significant hurdle for hate speech laws as it is difficult to draft one that does not exclude any particular group or point of view. If a hate speech law did manage to capture every possible group or viewpoint, it would have difficulty surmounting a challenge for overbreadth. While a review of cases and scholarly works reveals that hate speech laws seem to fail First Amendment scrutiny primarily because of vagueness, overbreadth, and content neutrality principles, such laws could also be voided if they fail to narrowly tailor a speech restriction.[108] In sum, the cumulative impact of the application of the above-described bedrock First Amendment principles has limited the scope of potential hate speech bans to expression that directly causes certain specific harm such as incitement to imminent violence, true threats, and harassment.

D. Observations

This review of First Amendment jurisprudence reveals several fundamental principles that shape the U.S. approach to the scope of freedom of expression. To begin, the First Amendment places the burden on the government to prove both content-based and content-neutral speech restrictions are valid, i.e., the burden is not on the speaker to demonstrate a right to speak.[109] Second, the First Amendment prohibits unduly vague and substantially overbroad bans on speech. This prohibition is a powerful check on the government's ability to regulate speech[110]—including hate speech.[111] Third, U.S. jurisprudence subjects both content-based and content-neutral restrictions on speech to heightened scrutiny, which means

107. *Id.* at 391–92.

108. Hate speech expert Nadine Strossen has assessed (consistent with the conclusions of human rights activists and civil society organizations in many countries) that non-censorial means are more effective in tackling the scourge of hate and intolerance than imposing penalties on speech, which can impact the narrow tailoring analysis for speech restrictions. *See* STROSSEN, *supra* note 3, at 133–82 (concluding that hate speech laws are ineffective and counterproductive in achieving tolerance and equality goals while highlighting the efficacy of a variety of non-censorial methods).

109. *See supra* notes 66–76 and accompanying text (discussing the government's burden under strict and intermediate scrutiny).

110. *See supra* notes 42–58 and accompanying text (discussing a variety of contexts in which speech bans fail due to vagueness and overbreadth).

111. *See supra* notes 99–102 and accompanying text (noting that the vagueness test is almost insurmountable in hate speech cases).

the restriction is presumed unconstitutional, and the government may overcome that presumption by proving the restriction is narrowly tailored to promote ends that are of compelling or substantial importance.[112] Fourth, although the Supreme Court has determined certain categories of speech are not worthy of First Amendment protections (e.g., incitement to imminent violence and true threats),[113] it has stated that even such speech may not be regulated in a way that discriminates based on content neutrality principles, which further limits the possibility of hate speech bans in the United States.[114] In sum, these interpretations make First Amendment jurisprudence highly speech-protective by forcing regulators to bear the burden of demonstrating restrictions are valid from a variety of angles.

II. THE U.N. APPROACH TO FREEDOM OF EXPRESSION

Part II analyzes the scope of freedom of expression in the U.N. human rights system. It begins by providing a brief background about the history and evolution of this global standard on speech, the U.N.'s enforcement machinery, and the differences between the U.N.'s approach and regional human rights standards. This Part then examines the key principles that form the foundation of the U.N.'s approach to freedom of expression. This analysis concludes with a focus on the U.N.'s treatment of hate speech.

A. Background

1. History and Evolution

The global standard for the protection of freedom of expression is set forth in Article 19 of the ICCPR,[115] which is one of the foundational treaties of the U.N. human rights system and a core

112. *See supra* notes 59–81 and accompanying text (examining strict scrutiny for content-based speech restrictions and intermediate scrutiny for content-neutral restrictions).

113. *See supra* notes 85–96 and accompanying text (describing categories of dangerous or disfavored, "low-value" speech that the Supreme Court has deemed unworthy of protection).

114. *See supra* notes 103–07 and accompanying text (discussing the application of content neutrality principles in the context of a case involving hateful and intolerant speech).

115. The International Covenant on Civil and Political Rights art. 19(2)–(3), *opened for signature* Dec. 16, 1966, S. Exec. Doc. E, 95-2, at 29 (1978), 999 U.N.T.S. 171, 178 (entered into force Mar. 23, 1976) [hereinafter ICCPR] (protecting speech of "all kinds" across frontiers and permitting restrictions only if provided by law and necessary to achieve public interest objectives).

component of the International Bill of Human Rights.[116] Negotiations to draft the ICCPR started soon after the United Nations was created and lasted about 20 years.[117] During the deliberations, there was a constant struggle between those nations fighting for broad protections for freedom of expression and those seeking numerous restrictions.[118] After the negotiations concluded in 1966, it took another ten years for the treaty to receive the minimum number of state ratifications to enter into force.[119] There are 173 states that are party to the ICCPR,[120] though implementation of the treaty's obligations has been uneven, including with respect to freedom of expression.[121]

The ICCPR neither created an international court to adjudicate treaty disputes nor referred such disputes to the International Court of Justice, as is the case in other human rights treaties.[122] Rather, the ICCPR created the U.N. Human Rights Committee ("HRC"), which is a group of independent experts elected by ICCPR State Parties to monitor the implementation of the treaty.[123] The HRC's main functions are to (1) make recommendations to each State Party about its record on fulfilling ICCPR obli-

116. PHILIP ALSTON & RYAN GOODMAN, INTERNATIONAL HUMAN RIGHTS 139 (2d ed. 2012) (noting the International Bill of Human Rights is comprised of the Universal Declaration of Human Rights, the ICCPR, and the International Covenant on Economic, Social, and Cultural Rights).

117. *Id.*

118. *See, e.g.*, Evelyn M. Aswad, *To Ban or Not to Ban Blasphemous Videos*, 44 GEO. J. INT'L L. 1313, 1320–22 (2013) (discussing the ICCPR's negotiating history involving advocacy by the USSR and its allies to include a mandatory ban on intolerant speech and resistance by the United States and its allies). The debates on freedom of expression that occurred with respect to the Universal Declaration of Human Rights, which was adopted in 1948, also displayed similar discussions. *See* Evelyn M. Aswad, *Losing the Freedom to Be Human*, 52 COLUM. HUM. RTS. L. REV. 306, 346–51 (2020) (describing attempts by the USSR and its allies to narrow freedom of expression provisions and refutations by other countries, including the United States).

119. ALSTON & GOODMAN, *supra* note 116, at 142.

120. *International Covenant on Civil and Political Rights*, UNITED NATIONS TREATY COLLECTION, https://bit.ly/36o25i3 [https://perma.cc/44GC-4EED] (last visited Aug. 8, 2021) [hereinafter ICCPR Treaty Collection].

121. *See infra* notes 154–206 and accompanying text (examining instances where the U.N. human rights machinery has found states did not implement ICCPR protections for speech).

122. *See, e.g.*, International Convention on the Elimination of All Forms of Racial Discrimination art. 22, *opened for signature* Mar. 7, 1966, S. Exec. Doc. C, 95-2 (1978), 660 U.N.T.S. 195, 218 (entered into force Jan. 4, 1969) [hereinafter CERD] (referring treaty disputes to the International Court of Justice); Convention on the Prevention and Punishment of the Crime of Genocide art. 9, *opened for signature* Dec. 9, 1948, 78 U.N.T.S. 277, 280 (entered into force Jan. 12, 1951) (designating the International Court of Justice as the dispute resolution forum).

123. ICCPR, *supra* note 115, at art. 28.

gations, (2) issue suggested interpretations of the treaty (known as "General Comments"), and (3) hear individual complaints lodged against a State Party that has consented to such a procedure.[124]

In addition, the U.N. system has an independent expert focused solely on free expression developments around the world: the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (the "Special Rapporteur").[125] The Special Rapporteur not only issues country-specific and thematic reports relating to free speech issues but also corresponds with governments about pending legislation, files amicus briefs in domestic litigation, and issues joint statements with other free expression experts.[126] When this Article refers to "U.N. standards" on freedom of expression, it is referring to U.N. treaties as interpreted by the relevant U.N. machinery, particularly the HRC and the Special Rapporteur.

The U.N. human rights machinery's views on freedom of expression have evolved over the last 45 years. In 1983, the HRC issued a short General Comment that consisted of a mere four paragraphs with recommendations on freedom of expression.[127] In 2011, the Committee issued General Comment 34, which provided detailed guidance on this human right and superseded the prior recommendations.[128] Emerging from an intensive, multi-year, and worldwide consultation process with State Parties and civil society,[129] General Comment 34 provided important clarifications and marked an evolution in the Committee's approach to freedom of expression by espousing broader protections for speech.[130] As this

124. *Id.* at art. 40; Optional Protocol to the International Covenant on Civil and Political Rights, *opened for signature* Dec. 16, 1966, art. 1, 999 U.N.T.S. 171, 302 (entered into force Mar. 23, 1976).

125. *Special Procedures of the Human Rights Council*, OHCHR, https://bit.ly/2TXYqoe [https://perma.cc/UK7J-U3TU] (last visited Aug. 8, 2021).

126. *Id.*

127. U.N. Hum. Rts. Comm., General Comment No. 10 (June 29, 1983), https://bit.ly/3hRZMZU [https://perma.cc/U285-DPAZ].

128. U.N. Hum. Rts. Comm., General Comment No. 34, U.N. Doc. CCPR/C/GC/34, ¶¶ 1–52 (Sept. 12, 2011) [hereinafter GC 34].

129. *See* Michael O'Flaherty, *Freedom of Expression: Article 19 of the International Covenant on Civil and Political Rights and the Human Rights Committee's General Comment 34*, 12 HUM. RTS. L. REV. 627, 650 (2012) (discussing contributions from a variety of stakeholders during the drafting process for General Comment 34).

130. *See id.* at 647–54 (commemorating the lead drafter of General Comment 34's description of how the HRC intentionally broadened its interpretations of freedom of expression, including with respect to access to information, a strengthening of governmental burdens to prove the validity of speech restrictions, and the approach to atrocity denial); Evelyn Mary Aswad, *To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?,* 77 WASH. & LEE L. REV.

Article compares the current state of U.N. standards on freedom of expression with current U.S. free speech law, the Article focuses on relevant U.N. developments from the last decade, i.e., since the adoption of General Comment 34.

2. Distinguishing U.N. Standards from Regional Standards

The distinction between U.N. and regional human rights standards is important because the UNGPs call for corporate respect of "internationally recognized human rights," which are defined as standards embodied in global U.N. instruments.[131] Given that U.N. norms serve as the common tapestry that binds U.N. member states, it is understandable that the UNGPs would be tied to U.N. (rather than regional) human rights standards. However, scholars and commentators often conflate U.N. and regional human rights systems when criticizing the U.N.'s approach to freedom of expression.[132] This section explains why it is inappropriate to conflate U.N. and regional standards, particularly when (1) comparing U.N. and First Amendment norms as well as (2) assessing whether companies should espouse the UNGPs' call to align content moderation with global standards.

Regional human rights instruments and monitoring mechanisms address the right to freedom of expression, but some provide for limitations beyond those commemorated in ICCPR Article 19's standard. For example, the Organization of Islamic Cooperation—the second largest international organization after the UN[133]—has developed a human rights system, which is based in part on the Cairo Declaration on Human Rights in Islam, an instrument that limits all rights, including free speech, to the principles set forth in Islam and national law.[134] In addition, the Human Rights Declara-

609, 637–42 (2020) (describing how the HRC interpretation of the scope of freedom of expression expanded with the adoption of General Comment 34).

131. UNGPs, *supra* note 11, at Principle 12. The UNGPs define "internationally recognized human rights" as constituting at a minimum the U.N.'s International Bill of Human Rights (which includes the ICCPR) as well as an International Labor Organization Declaration. *Id.* The commentary notes that additional U.N. instruments may be referred to as well. *Id.* at commentary to Principle 12.

132. *See* Aswad, *supra* note 130, at 614–43 (describing how academics and practitioners have collapsed U.N. and regional standards when finding fault with U.N. standards).

133. *Member States*, ORG. OF ISLAMIC COOPERATION, https://bit.ly/36sEfS1 [https://perma.cc/3F5T-RL6V] (last visited Aug. 8, 2021) (noting the organization has 57 member states).

134. Org. of Islamic Cooperation, *The Cairo Declaration on Human Rights in Islam*, art. 25, adopted Nov. 28, 2020. https://bit.ly/2TLTL99 [https://perma.cc/CB34-JXRV] [hereinafter *Cairo Declaration*]. The Cairo Declaration also specifies

tion of the Association of South East Asian Nations[135] contains text that curtails rights, including free expression, in a variety of ways that are inconsistent with U.N. standards.[136]

Other regional human rights instruments contain freedom of expression protections that have phrasing similar to ICCPR Article 19, but the relevant monitoring machinery has interpreted those instruments in a more restrictive way, as is the case with the European Convention on Human Rights ("ECHR").[137] For example, whereas the HRC has taken the position that blasphemy and the denial of historic atrocities are protected under the ICCPR,[138] the European Court of Human Rights ("ECtHR") has upheld criminal sanctions for both.[139] The ECtHR often refuses to consider hate

---

that "[f]reedom of expression should not be used for denigration of religions and prophets or to violate the sanctities of religious symbols or to undermine the moral and ethical values of society." *Id.* at art. 21(c). The text of the Declaration's freedom of expression provision does not require that speech restrictions be necessary or proportional to the interests at stake. *See id.* at art. 21(a)–(c). Such approaches conflict with the UN's protections for freedom of expression. See *infra* note 182 and accompanying text (explaining that the protection of religion is not a legitimate reason to restrict speech under U.N. standards) and note 187 (observing that any restrictions on speech must be justified as the least intrusive means to achieve legitimate objectives and must be proportional to the interests at stake). However, the Cairo Declaration also contains a clause to the effect that nothing in the declaration may undermine the obligations of states under international human rights treaties. *Cairo Declaration*, at art. 25(b).

135. *ASEAN Human Rights Declaration*, ASS'N OF SE. ASIAN NATIONS [ASEAN] (Nov. 19, 2012), https://bit.ly/2TOf4XE [https://perma.cc/7PW7-UQFC].

136. The Declaration's curtailments of universally recognized rights include:

> the use of the concept of 'cultural relativism' to suggest that rights in the [Universal Declaration on Human Rights] do not apply everywhere; stipulating that domestic laws can trump universal human rights; incomplete descriptions of rights that are memorialized elsewhere; introducing novel limits to rights; and language that could be read to suggest that individual rights are subject to group veto.

Press Release, U.S. Dep't of State, ASEAN Declaration on Human Rights Press Statement (Nov. 20, 2012), https://bit.ly/2UwNe1Y [https://perma.cc/3KS3-V2LY].

137. The European Convention on Human Rights protects the right "to receive and impart information and ideas without interference by public authority and regardless of frontiers" and allows for restrictions that are (1) prescribed by law and (2) "necessary in a democratic society" (3) to achieve a legitimate public interest objective. European Convention for the Protection of Human Rights and Fundamental Freedoms art. 10, *opened for signature* Nov. 4, 1950, Euro. T. S. No. 5, 213 U.N.T.S. 221 (entered into force Sept. 3, 1953). To compare ECHR Article 9 with the ICCPR's protection for expression, see *infra* notes 145–47 and accompanying text.

138. *See* GC 34, *supra* note 128, at ¶¶ 48–49.

139. *See, e.g.*, *Otto-Preminger-Institute*, 295 Eur. Ct. H.R. (ser. A), ¶¶ 51–57 (1994); Garaudy v. France, 2003-IX Eur. Ct. H.R. 369. The OIC's Independent Permanent Human Rights Commission has applauded such ECtHR caselaw. Press Release, OIC Independent Permanent Human Rights Commission, (Dec. 10, 2020), https://bit.ly/2UwNrlM [https://perma.cc/R2MK-2P3M] (highlighting "the

speech claims because it believes they fall outside the ECHR's scope of protection of ECHR, but the U.N. machinery requires governments to prove restrictions are lawful, even for the most heinous hate speech.[140] Moreover, the U.N. machinery requires governments to prove that any speech restriction constitutes "the least intrusive means" to achieve a legitimate public interest objective.[141] But the ECtHR has used a more lenient standard in determining whether a speech limitation is "necessary" to achieve a governmental objective.[142]

While it is indeed unfortunate when *regional* human rights systems provide fewer protections than the *U.N.* system (and when countries seek to justify their breaches of U.N. treaties by invoking such regional norms), it is inappropriate to conflate U.N. treaties with regional treaties. As a matter of established international treaty law, the fact that regional treaties grant differing levels of protection from U.N. treaties does not undermine or change the scope of U.N. treaty obligations. Under the Vienna Convention on the Law of Treaties, the only other treaties that can be used to interpret the ICCPR are those that, among other things, all the

---

jurisprudence emerging from the European Court of Human Rights, which validates restrictions on freedom of expression criticizing religious beliefs where such expression constitutes incitement to hatred and is deemed offensive to the adherents of a particular religion.").

140. *See* David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 26, U.N. Doc. A/74/486 (Oct. 9, 2019), [hereinafter *Special Rapporteur 2019 Report*] (comparing U.N. and European protections for freedom of expression).

141. GC 34, *supra* note 128, at ¶ 34 (stating that limitations on speech "must be the least intrusive instrument amongst those which might achieve their protective function").

142. *See* JONAS CHRISTOFFERSEN, FAIR BALANCE: PROPORTIONALITY, SUBSIDIARITY AND PRIMARITY IN THE EUROPEAN CONVENTION ON HUMAN RIGHTS 129 (2009) (observing the ECtHR's "general rejection of the least/less onerous means-test" in assessing violations of freedom of expression). The U.N. machinery and the ECtHR have developed additional different rules of interpretation relating to their respective treaties, which further explains their divergent outcomes in similar cases. For example, the ECtHR applies a "margin of appreciation" in which it defers to governmental judgments in deciding cases. ALSTON & GOODMAN, *supra* note 116, at 946–48 (describing how the ECtHR has adopted a practice of deferring to governments, particularly when treaty parties display a varied practice in implementing a particular right). But the HRC has explicitly rejected granting such deference to governmental authorities. GC 34, *supra* note 128, at ¶ 36 ("[T]he Committee recalls that the scope of this freedom is not to be assessed by reference to a 'margin of appreciation' . . . [rather] a State party . . . must demonstrate in specific fashion the precise nature of the threat . . . that has caused it to restrict freedom of expression.").

ICCPR parties have endorsed,[143] which is not the case with respect to treaties that encompass states from a particular region.

This is not to say that regional treaties and their jurisprudence may not be useful to examine in considering potential applications of the ICCPR, but it does mean that differing regional interpretations do not define ICCPR standards or somehow undermine the coherency of U.N. treaty standards. Indeed, if international treaty law allowed a subgroup of states to redefine the scope of U.N. treaty standards through regional arrangements, that would create enormous instability in multilateral treaty relations and undermine incentives to join global treaties. Instead, as noted by the Special Rapporteur, when regional systems provide fewer protections for expression than the U.N. system, countries that follow those regional standards are violating their U.N. treaty obligations.[144]

3. Overview

ICCPR Article 19 broadly defines freedom of expression as the "freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."[145] The treaty permits (but does not require) governments to limit speech when they can prove that each part of a tripartite test is met.[146] Any speech restrictions must be (1) "provided by law," (2) imposed to achieve one of the treaty's enumerated public interest objectives, and (3) "necessary" to achieve those objectives.[147] These three

143. Vienna Convention on the Law of Treaties art. 31(2)–(3), *opened for signature* May 23, 1969, 1155 U.N.T.S. 331, at 340 (entered into force Jan. 27, 1980) [hereinafter Vienna Convention]. Under the Vienna Convention, other treaties or instruments may be used as part of the context to interpret a convention when there is (1) an "agreement relating to the treaty which was made between *all* the parties in connection with the conclusion of the treaty," (2) an "instrument which was made by one or more parties in connection with the conclusion of the treaty and *accepted by the other parties* as an instrument related to the treaty," or (3) a "subsequent agreement *between the parties* regarding the interpretation of the treaty or the application of its provisions." *Id.* (emphasis added). In addition, State Party violations of treaty obligations do not, as a legal matter, negate the scope or coherency of those obligations. Under the Vienna Convention, subsequent state practice may only be used to interpret the scope of treaty obligations when that practice "establishes the agreement of the parties regarding its interpretation." *Id.* at art. 31(3)(b).

144. *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 26 (noting that "*Regional* human rights norms cannot, in any event, be invoked to justify departure from *international* human rights protections") (emphasis added).

145. ICCPR, *supra* note 115, at art. 19(2).

146. *Id.* at art. 19(3).

147. *Id.* (providing that freedom of expression may "be subject to certain restrictions, but these shall only be such as are *provided by law* and are *necessary*: (a)

prongs are commonly referred to as the legality, legitimacy, and necessity conditions.[148] If a State Party fails in its burden of demonstrating that each part of the tripartite test is met, then the speech restriction is invalid under the ICCPR.[149] In addition, any restrictions on speech should respect the other rights in the ICCPR, including a ban on discrimination.[150] The next Section examines U.N. interpretations of ICCPR Article 19(3)'s tripartite test of legality, legitimacy, and necessity since the 2011 adoption of the HRC's pivotal interpretations in General Comment 34.

B. Key Principles

1. The Legality Test

The U.N. human rights machinery has identified multiple components of ICCPR Article 19(3)'s legality (or "provided by law") condition for imposing restrictions on speech, including that laws regulating speech must not be impermissibly vague, over broad, or improperly enacted. The HRC has stated that, to avoid being impermissibly vague, speech restrictions (1) "must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly," (2) must be "made accessible to the public," and (3) "may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution."[151] The Special Rapporteur has reinforced this formulation and emphasized the danger of chilling speech.[152]

The U.N. machinery has concluded that a variety of speech restrictions imposed by both authoritarian and democratic regimes in every geographic region contain inappropriately vague bans on

---

For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals") (emphasis added).

148. *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 8.

149. *See* GC 34, *supra* note 128, at ¶¶ 22, 27.

150. *Id*. at ¶ 26 (noting speech restrictions must not only comply with ICCPR Article 19 but also the other provisions of the ICCPR, including "the non-discrimination provisions of the Covenant").

151. GC 34, *supra* note 128, at ¶ 25.

152. *See, e.g.*, *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 8; David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Affidavit for Third-Party Intervention Filed Before ECOWAS Court*, at ¶ 13 (May 18, 2016), https://bit.ly/3wxdV3Z [https://perma.cc/97NF-6HW6] [hereinafter *Special Rapporteur Affidavit for ECOWAS Court*] (noting that vague speech restrictions "create a 'chilling effect' that discourages individuals from exercising their rights to free expression for fear that government authorities may use their broad interpretive discretion to penalize a swath of speech-related activities").

speech.[153] For example, the U.N. Special Rapporteur has observed that a number of bans on seditious speech contain vague language that does not meet the legality condition.[154] U.N. experts have also noted that various limitations on criticism of government officials[155]

153. The examples of terminology set forth in *infra* notes 154–67 focus primarily on vagueness concerns, but in some instances the U.N. machinery has also identified overbreadth problems. For the topic of overly broad speech restrictions, see *infra* notes 168–71 and accompanying text.

154. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. MYS 6/2018, at 3 (Dec. 28, 2018), https://bit.ly/2T0Uh2t [https://perma.cc/LX3J-V43Y] [hereinafter *2018 Special Rapporteur Comment on Malaysia*] (condemning as unduly vague phrasing that outlawed acts with a "'seditious tendency,' including any act that conjures feelings of 'hatred,' 'contempt,' 'disaffection,' 'discontent,' 'ill will,' or 'hostility'"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. OMN 1/2018, at 3 (Mar. 26, 2018), https://bit.ly/3AOTpyT [https://perma.cc/CJ3Q-PBVR] (finding Oman's legal bar on speech that "prejudices the independence, unity or territorial integrity of the country" to be inappropriately vague); *Special Rapporteur Affidavit for ECOWAS Court*, *supra* note 152, at 15 (noting Gambia's sedition offenses contain vague provisions, such as "an intention 'to raise discontent or disaffection among the inhabitants of the Gambia' or 'to promote feelings of ill will and hostility between different classes of the population'"); Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on His Mission to Israel and the Occupied Palestinian Territory*, ¶¶ 28–30, A/HRC/20/17/Add.2 (June 11, 2012), [hereinafter *Special Rapporteur Views on Israel and the Occupied Palestinian Territory*] (finding a law that penalizes questioning the existence of Israel to be unduly vague); Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 584, U.N. Doc. A/HRC/17/27/Add.1 (May 27, 2011) [hereinafter *Addendum to 2011 Special Rapporteur Report to Human Rights Council*] (criticizing China for a vague speech prohibition on the "subversion of state power").

155. *See, e.g.*, Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 29, U.N. Doc. A/66/290 (Aug. 10, 2011) [hereinafter *2011 Special Rapporteur Report to UNGA*] (condemning as unduly vague a prohibition on "instigating hatred and disrespect against the ruling regime"); *Addendum to 2011 Special Rapporteur Report to Human Rights Council*, *supra* note 154, at ¶ 1173 (criticizing as vague Iran's penal code for banning "propaganda against the system," "insulting of leaders of the country," "insults against the President of the Islamic Republic of Iran," and "establishment of anti-revolutionary media"); Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 20, A/HRC/20/17 (June 4, 2012) (criticizing the vagueness of a law that protects the monarchy from criticism and insult).

and speech that could result in unrest[156] are improperly vague. On numerous occasions, U.N. experts have found that speech prohibitions imposed to protect national security contain inappropriately vague language.[157]

The U.N. machinery has identified a worldwide trend involving improperly vague speech restrictions used to fight terrorism, extremism, and radicalization. For example, U.N. experts have repeatedly condemned as vague a variety of country-specific laws banning the "glorification," "promotion," or "justification" of "terrorism."[158] The HRC has also issued a general call to the international

156. *See, e.g.*, *2011 Special Rapporteur Report to UNGA*, *supra* note 155, at ¶¶ 26, 29 (condemning as vague bans on "offenses that damage public tranquility"); *Special Rapporteur Affidavit for ECOWAS Court*, *supra* note 152, at 15 (criticizing as inappropriately vague a ban on speech that promotes "feelings of ill will and hostility between different classes of the population"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. BGD 4/2018, at 4 (May 14, 2018), https://bit.ly/3hsW6P1 [https://perma.cc/9XXU-36XZ] [hereinafter *2018 Special Rapporteur Comment on BGD*] (criticizing a draft Bangladeshi bill that would prohibit speech which "ruins communal harmony or creates instability or disorder or disturbs or is about to disturb the law and order situation").

157. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. JOR 3/2018, at 3 (Dec. 7, 2018), https://bit.ly/3huVMzv [https://perma.cc/V97G-LDFJ] [hereinafter *Special Rapporteur Views on Jordan*] (finding that a penal code's prohibition of speech that "would subject Jordan to the danger of violent acts or disturb its relations with a foreign state" did not "meet the level of clarity and predictability as required by international human rights law"); Frank La Rue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. HUN 2/2012, at 5 (Mar. 14, 2012), https://bit.ly/3htQG6w [https://perma.cc/QH68-A5YL] (determining that a Hungarian law that could be used to compel disclosure of journalistic sources in the "interest of protecting national security and public order" was unduly vague); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. CHN 7/2015, at 4–5 (Aug. 5, 2015), https://bit.ly/2T06ILY [https://perma.cc/Z232-FXNA] [hereinafter *Special Rapporteur Views on China*] (criticizing Chinese cybersecurity draft legislation for vagueness in banning network activity that could be construed as "harming national security").

158. *See, e.g.*, Off. of the High Comm'r for Hum. Rts., Joint Commc'n by U.N. Special Procedures to the Gov't of France No. FRA 1/2015, at 6 (Feb. 2, 2015), https://bit.ly/2UASOk3 [https://perma.cc/5FBR-WNHF] (translated text) (calling on France to avoid speech bans that use vague phrasing such as "glorification" or "promotion" of terrorism); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on His Mission to Turkey*, ¶ 84, A/HRC/35/22/Add.3 (June 21, 2017) [hereinafter *Special Rapporteur Report on Turkey*] (urging Turkey to define crimes such as "encouragement of terrorism," "extremist activity," and "praising," "glorifying," or "justifying" terrorism); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Joint Commc'n No. ESP 3/2015, at 10 (Feb. 17, 2015), https://bit.ly/

community to avoid the use of such vague terminology in the counter-terrorism context.[159] Similarly, the U.N. Special Rapporteur has condemned social media speech codes that duplicate such bans to combat extremism and terrorism on their platforms.[160]

The U.N. human rights machinery has repeatedly criticized as unduly vague laws that ban "unfounded," "biased," "false," or "fake" information,[161] including where expression is likely to

---

3wxefjd [https://perma.cc/CY6R-83EA] [hereinafter *Joint Communication on Spanish Legislation*] (translated text) (criticizing bans on "glorifying or promoting terrorism"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on His Visit to Tajikistan*, ¶¶ 39, 72, U.N. Doc. A/HRC/35/22/Add.2 (Oct. 13, 2017) [hereinafter *Special Rapporteur Report on Tajikistan*] (critiquing "vague definitions of what constitutes 'extremism' and 'terrorism'").

159. *See, e.g.*, GC 34, *supra* note 128, at ¶ 46 ("Such offences as 'encouragement of terrorism' . . . 'extremist activity' . . . 'praising', 'glorifying,' or 'justifying' terrorism, should be clearly defined . . . ."); Off. of the High Comm'r for Hum. Rts., *Joint Declaration on Freedom of Expression and Countering Violent Extremism* (May 4, 2016), https://bit.ly/3r81nPu [https://perma.cc/VJ4P-2TW2] (noting that governmental initiatives often have unclear definitions of "extremism" and "radicalization"); Off. of the High Comm'r for Hum. Rts., *Joint Declaration on Freedom of Expression and Responses to Conflict Situations* (May 4, 2015), https://bit.ly/3hY6MEi [https://perma.cc/PBS8-LWEH] [hereinafter *Joint Declaration on Responses to Conflict Situations*] (calling on states to avoid vague concepts "such as glorifying', 'justifying' or 'encouraging' terrorism").

160. *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 26 ("Company prohibitions of threatening or promoting terrorism, supporting or praising leaders of dangerous organizations and content that promotes terrorist acts or incites violence are, like counter-terrorism legislation, excessively vague.").

161. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Research Paper 1/2019 on Freedom of Expression and Elections in the Digital Age*, at 9 (June 2019), https://bit.ly/3APxV52 [https://perma.cc/E8XW-YDDZ] [hereinafter *Elections in the Digital Age*].

> [V]ague and highly subjective terms—such as 'unfounded,' 'biased,' 'false,' and 'fake'—do not adequately describe the content that is prohibited. As a result, they provide the authorities with broad remit to censor the expression of unpopular, controversial or minority opinions, as well as criticism of the government and politicians in the media and during electoral campaigns.

Id.; Off. of the High Comm'r for Hum. Rts., *Joint Declaration on Freedom of Expression and Elections in the Digital Age*, ¶ 1.a.iii (Apr. 30, 2020), https://bit.ly/3wA6J6V [https://perma.cc/N3ZN-SFCC] [hereinafter *Joint Elections Declaration*] ("There should be no general or ambiguous laws on disinformation, such as prohibitions on spreading 'falsehoods' or 'non-objective information.'"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. ITA 1/2018, at 4 (Mar. 20, 2018), https://bit.ly/3htdL9m [https://perma.cc/5SK3-WLPR] (criticizing Italian restrictions on "manifestly unfounded and biased news" as those terms "are not defined and therefore raise concerns of vagueness"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. FRA 5/2018, at 6 (May 28,

"cause fear or alarm."[162] The Special Rapporteur has also expressed vagueness concerns about laws that regulate discussion of historic events.[163] In addition, U.N. experts have called for bans on defamation or slander to define those terms with precision.[164]

U.N. experts have also identified improperly vague laws that infringe on speech in other contexts. For example, the U.N. machinery has found that laws requiring respect for decency/cultural norms trigger vagueness issues.[165] Similarly, laws that mandate respect for religious sensibilities have been deemed impermissibly

2018), https://bit.ly/2SYXSy0 [https://perma.cc/29YT-4NS8] (translated text) (expressing vagueness concerns about a French bill that would prohibit "false information likely to affect the fairness of the ballot").

162. *Special Rapporteur Affidavit for ECOWAS Court*, *supra* note 152, at 15 (determining that the criminalization of statements "likely to cause fear and alarm to the public, when the journalist has reason to believe the report is false" contains inappropriately vague terms that will cause uncertainty among journalists).

163. *See Special Rapporteur Views on Israel and the Occupied Palestinian Territory*, *supra* note 154, at ¶¶ 28–30 (noting the vagueness of a law that restricts expression of mischaracterizations of historic events).

164. *See, e.g.*, *Special Rapporteur Views on China*, *supra* note 157, at 4–5 (expressing concern that a draft cybersecurity law was vague with respect to a number of terms, including slander and defamation).

165. *See, e.g.*, *Special Rapporteur Views on Israel and the Occupied Palestinian Territory*, *supra* note 154, at ¶ 51 (June 11, 2012) (criticizing a vague ban on speech that "contradicts principles of freedom, national responsibility or [is] 'inconsistent with morals' "); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. LAO 2/2016, at 3 (May 6, 2016), https://bit.ly/2T0UB1b [https://perma.cc/Q6EG-J9AC] (expressing vagueness concerns about a law in Laos that requires foreign media and others to abide by local traditions and culture); Off. of the High Comm'r for Hum. Rts., *Joint Declaration on Media Independence and Diversity in the Digital Age*, at 4 (2018), https://bit.ly/2SYXVKc [https://perma.cc/KXS4-QB9A] [hereinafter *Joint Declaration on Media Independence*] (stating that "cultural security" is an "inherently vague notion" and should not be used to restrict freedom of expression).

vague.[166] Laws requiring neutrality by the media and academics have also failed this test.[167]

The legality test also includes a requirement that a speech restriction not be overly broad.[168] While speech bans can often fail the legality test for both vagueness and overbreadth, U.N. experts have found a variety of speech prohibitions to violate the overbreadth principle in particular. For example, restrictions that aim to counter terrorism or cybercrime are often phrased in overly broad terms.[169] Restrictions that seek to prevent the circulation of offen-

166. *See, e.g.*, *2011 Special Rapporteur Report to UNGA*, *supra* note 155, at ¶¶ 26, 29 (condemning as unduly vague phrasing that prohibits "incitement to religious unrest," "promoting division between religious believers and non-believers," and "defamation of religion"); *Addendum to 2011 Special Rapporteur Report to Human Rights Council*, *supra* note 154, at ¶ 1173 (criticizing as vague Iranian bans on speech that includes "enmity against God," "insulting Islamic sanctities," and "distribution of pictures and materials intended to mock sanctities"); *2018 Special Rapporteur Comment on BGD*, *supra* note 156, at 4 (noting the criminalization of speech that "injures religious feelings" fails the vagueness test); *2018 Special Rapporteur Comment on Malaysia*, *supra* note 154, at 3 (criticizing a vague prohibition on speech that "promote[s] feelings of ill will, hostility or hatred . . . on the ground of religion"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. MDV 1/2020, at 4–5 (May 19, 2020), https://bit.ly/3hTEFGD [https://perma.cc/P3G7-83L2] [hereinafter *Special Rapporteur Views on the Maldives*] (criticizing a law that bans speech which engenders "religious discord amongst people").

167. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteurs Commc'n No. EGY 1/2014, at 6 (Jan. 14, 2014), https://bit.ly/36nGMgu [https://perma.cc/8TBL-R7CE] (expressing concern that an Egyptian law mandating neutrality for the press contained "unclear language and the possible violations of the freedom of the press . . . given the vaguely defined role of the State in the oversight of the 'neutrality' of all media"); Koumbou Boly Barry (Special Rapporteur on the Right to Education), Joint Special Rapporteurs Commc'n No. BRA 4/2017, at 5, 7 (Apr. 13, 2017), https://bit.ly/3xwMsAJ [https://perma.cc/VD3C-G2V6] [hereinafter *Joint Special Rapporteur Views on Brazil*] (highlighting vagueness of draft bills that require teachers to present ideas "in a fair manner" and to refrain from "political partisan propaganda").

168. *Affidavit for Third-Party Intervention Filed Before ECOWAS Court*, *supra* note 152, at 12 (noting that restrictions "must not be overly broad or vague . . . . Legal restrictions on expression that are too broad or vague create a 'chilling effect' "); *Joint Declaration on Responses to Conflict Situations*, *supra* note 159, at ¶ 3 (calling for speech restrictions to "conform strictly to international standards, including by . . . not employing vague or unduly broad terms").

169. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. GBR 6/2018, at 2 (July 17, 2018), https://bit.ly/36tmimC [https://perma.cc/K53A-P565] (critiquing overbroad wording in a draft UK bill that would make it an offense "to express support of a proscribed organization" where there is no link to an intention to cause harm, particularly as such wording could even "apply to the activities of human rights organizations and associations, including those providing legal opinions defending the rights of members of a proscribed

sive speech have similarly failed this test.[170] In addition, laws that purport to avert a variety of dangers or maintain public order have also contravened this principle.[171]

U.N. experts have not interpreted the legality test as solely constituting a prohibition on vague or overbroad speech restrictions. The U.N. machinery has also interpreted "provided by law" in ICCPR Article 19 to mean restrictions must be properly enacted, which encompasses a variety of good governance/rule of law components. For example, states should follow regular procedures for law-making,[172] and should subject proposed restrictions to public

---

organization"); *Id.* (criticizing as overly broad phrasing in a UK draft bill that would criminalize "the publication of an image of an item of clothing or 'any other article' in such a way or circumstances as to arouse 'reasonable suspicion' that a person is member or supporter of a proscribed organization" particularly given the lack of a link to incitement to violence); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. PAK 13/2015, at 2–5 (Dec. 14, 2015), https://bit.ly/36ve9Ot [https://perma.cc/A8DR-YW4U] (criticizing a Pakistani cybercrime law as overbroad because the phrasing "effectively criminalizes the accessing, copying and transmitting of any information system or data").

170. *Addendum to 2011 Special Rapporteur Report to Human Rights Council*, *supra* note 154, at ¶¶ 939–40 (finding Hungary's ban on "commercial communication that 'may not infringe upon human dignity'" to be improperly broad); *Id*. at ¶¶ 935–36 (finding a requirement that "viewers or listeners shall be given a forewarning prior to broadcasting any image or sound effects that may potentially infringe a person's religion, faith-related or other philosophical convictions or which are violent or otherwise disturbing" to be improperly broad); *Id*. at ¶ 941 (determining that a prohibition on "commercial communication broadcasted in the media service" from conveying "'religious, conscientious or philosophical convictions except for commercial communications broadcasted in thematic media services concerning a religious topic" to be unduly broad).

171. *Id.* at ¶ 937 (finding the phrases "state of distress," "state of emergency" and "state of extreme danger" to be overly broad in a law requiring media to carry certain messaging during such times). Often public order justifications for speech restrictions trigger both vagueness and overbreadth concerns. *See* Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 52, U.N. Doc. A/67/357 (Sept. 7, 2012) [hereinafter *Special Rapporteur 2012 UNGA Report*] (noting as vague and overly broad the following prohibitions: "expression of feelings of hostility," "outraging religious feelings," "provocation of sectarian or racial division," "inciting unlawful acts," and "'inciting people to disputes").

172. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 37, U.N. Doc. A/72/350 (Aug. 18, 2017) ("The requirement of legality ("provided by law") requires that regular procedures be followed in the adoption of restrictions . . . ."); *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 7 (noting that the legality test requires laws be adopted "by regular legal processes" and that "[s]ecretly adopted restrictions fail this fundamental requirement"); *Elections in the Digital Age*, *supra* note 161, at 11–12 (noting that the "covert and illicit nature [of Distributed Denial of Service] attacks usually violate the requirement

comment prior to adoption.[173] Moreover, the Special Rapporteur has repeatedly called for infringements on expression to be subject to judicial review as part of the "provided by law" test.[174]

2. The Legitimacy Test

The text of ICCPR Article 19(3) provides that speech may only be restricted for certain public interest purposes: "(a) [f]or respect of the rights or reputations of others" or "(b) [f]or the protection of

---

that restrictions on freedom of expression must be 'provided by law'"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. HUN 1/2017, at 3 (Apr. 1, 2017), https://bit.ly/3k3W0zo [https://perma.cc/T2G3-6V2U] ("Under [domestic law], an impact assessment must be carried out before the adoption of legislation. . . . [N]o impact assessment was made [in this case]. The lack of consultations and Parliamentary negotiation therefore appear to undermine any argument that the law's restrictions are 'provided by law.'"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. USA 4/2020, at 6 (Mar. 19, 2020), https://bit.ly/3k4UX26 [https://perma.cc/83SH-G4YE] (raising concerns about the legality test when a U.S. bill would set up a rule making methodology that departs from regular procedures).

173. *See, e.g.*, David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 32, U.N. Doc. A/HRC/29/32 (May 22, 2015) [hereinafter *2015 Special Rapporteur Report to Human Rights Council*] ("Proposals to impose restrictions on encryption or anonymity should be subject to public comment and adopted, if at all, according to regular legislative process."); *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 6 ("Rules should be subject to public comment and regular legislative or administrative processes."); *Special Rapporteur Report on Tajikistan*, *supra* note 158, at ¶ 80 ("Private sector representatives and civil society must be consulted and included in the promotion of a new regulatory system."); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 15, U.N. Doc. A/71/373 (Sept. 6, 2016) [hereinafter *Special Rapporteur 2016 Report to UNGA*] (noting concerns about the lack of public engagement prior to adoption of laws affecting expression in Montenegro, Brazil, and Russia).

174. *See, e.g.*, *Joint Declaration on Media Independence*, *supra* note 165, at ¶ 3(f) (stating that restrictions on expression should "be subject to judicial oversight"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Procedures Commc'n No. CHN 21/2018, at 3 (Nov. 12, 2018), https://bit.ly/3qYiovr [https://perma.cc/KD5K-7MAE] [hereinafter *Special Rapporteur 2018 Views on China*] ("Restrictions must meet the standards of legality, meaning that they are publicly provided by a law which meets standards of clarity and precision, and are interpreted by independent judicial authorities."); *2015 Special Rapporteur Report to Human Rights Council*, *supra* note 173, at ¶ 32 (stating that "a court, tribunal or other independent adjudicatory body must supervise the application of the restriction" on encryption or anonymity to be "provided by law"); *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 7 ("The assurance of legality should generally involve the oversight of independent judicial authorities.").

national security or of public order (ordre public), or of public health or morals."[175] In interpreting this "legitimacy" condition, the U.N. human rights machinery has emphasized that the enumerated list of public purposes is a limited list which governments should interpret narrowly.[176]

U.N. experts have provided guidance on the legitimacy of various public interest justifications. For example, U.N. experts have warned that invocations of national security are often pretexts for suppressing lawful speech[177] or are deployed in an otherwise improper manner.[178] With respect to the "rights of others," the HRC

175. ICCPR, *supra* note 115, art. 19(3).

176. *See* GC 34, *supra* note 128, at ¶ 22 ("Restrictions are not allowed on grounds not specified in paragraph 3, even if such grounds would justify restrictions to other rights protected in the Covenant."); *2015 Special Rapporteur Report to Human Rights Council*, *supra* note 173, at ¶ 33 (observing that speech limitations may only be imposed for reasons specified in the ICCPR and noting "[n]o other grounds may justify restrictions on freedom of expression. Moreover, because legitimate objectives are often cited as a pretext for illegitimate purposes, the restrictions themselves must be applied narrowly."); *see also Addendum to 2011 Special Rapporteur Report to Human Rights Council*, *supra* note 154, at ¶ 892 (highlighting that the ICCPR has fewer permissible grounds than the ECHR, which includes the interest of territorial integrity and other purposes, but noting that "[a]lthough these [ECHR] purposes . . . may be taken into account on a case-by-case basis, a relatively limited number of reasons for permissible interference in the ICCPR indicated that 'these are to be interpreted narrowly in cases of doubt'").

177. GC 34, *supra* note 128, at ¶ 30 ("It is not compatible with [ICCPR Article 19(3)] to invoke [national security] laws to suppress or withhold from the public information of legitimate public interest that does not harm national security or to prosecute journalists, researchers, environmental activists, human rights defenders, or others for having disseminated such information."). Additionally,

> [A] restriction sought to be justified on the ground of national security is not legitimate if its genuine purpose or demonstrable effect is to protect interests unrelated to national security, including, for example, to protect Government from embarrassment or exposure of wrongdoing, or to conceal information about the functioning of its public institutions, or to entrench a particular ideology or to suppress industrial unrest.

Addendum to 2011 Special Rapporteur Report to Human Rights Council, *supra* note 154, ¶ 584. *Special Rapporteur 2016 Report to UNGA*, *supra* note 173, at ¶ 18 (explaining that national security "should be limited in application to situations in which the interest of the whole nation is at stake, which would thereby exclude restrictions in which the interest of a Government, regime or power group [are protected]").

178. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Third Party Intervention filed in the European Court of Human Rights*, at ¶ 20 (June 3, 2019), https://bit.ly/3r2j6aJ [https://perma.cc/696F-DVB3].

> States regularly invoke national security to legitimise surveillance measures that entail over-broad restrictions on human rights. The invocation of national security does not in and of itself provide an adequate human rights law justification. Rather, the State must provide an '*articulable and*

has noted that "'rights' includes human rights as recognized in the [ICCPR] and more generally in international human rights law."[179] The HRC has stated that restrictions to protect "morals" should not be assessed from "'principles . . . deriving exclusively from a single tradition.' Any such limitations must be understood in the light of the universality of human rights and the principle of non-discrimination."[180] With regard to "public order," the U.N. machinery has noted that such invocations "must be limited to specific situations in which a limitation would be demonstrably warranted."[181]

In applying such parameters, the U.N. machinery has noted that a variety of purposes plainly do not serve legitimate public interest objectives. For example, laws that "protect religions against criticism or prohibit the expression of dissenting religious beliefs" are not based on a legitimate public interest objective.[182] In addi-

---

evidence-based justification for the interference'. The State must, at a minimum, give a meaningful public account of the tangible benefits.

Id. (emphasis added). *Joint Declaration on Freedom of Expression and the Internet*, ¶¶ 6.b, 6.d (June 1, 2011), https://bit.ly/3e69ir0 [https://perma.cc/AL2W-Z63U] [hereinafter *Joint Declaration on Expression and the Internet*] ("Cutting off access to the Internet, or parts of the Internet, for whole populations or segments of the public . . . can never be justified, including on public order or national security grounds. The same applies to slow-downs imposed on the Internet or parts of the Internet."); David Kaye, (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. USA 2/2019, at 2–4 (Feb. 14, 2019), https://bit.ly/3htdVgY [https://perma.cc/XYE2-9GXD] (noting that proposed anti-boycott legislation could not be justified by an invocation of national security or public order despite the U.S. government having "a legitimate interest in standing with other governments politically").

179. GC 34, *supra* note 128, at ¶ 28. The HRC provided as an example that, while it may be acceptable to limit speech to protect the right to vote in instances of voter intimidation or coercion, "such restrictions must not impede political debate." *Id.*; *see also Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 7 (noting that the rights of others, as defined by the HRC, would include "rights to privacy, life, due process, association and participation in public affairs, to name a few").

180. GC 34, *supra* note 128, at ¶ 32; *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 7 (noting the HRC's cautions on the appropriate interpretation of the protection of morals); *see also* David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Special Rapporteur Commc'n No. PSE 2/2017, at 3 (Aug. 16, 2017), https://bit.ly/3xw5qap [https://perma.cc/5J5M-8RHG] (critiquing a Palestinian cybercrime law that criminalizes "infringing upon public morals" because the "law provides no guidance on what is deemed to disrupt or go against public order or morals").

181. *Special Rapporteur 2016 Report to UNGA*, *supra* note 173, at ¶ 18.

182. GC 34, *supra* note 128, at ¶ 48 (stating that prohibitions on speech based solely on a "lack of respect for religion or other belief system, including blasphemy laws, are incompatible with" the ICCPR); *see Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 21 ("[A]nti-blasphemy laws fail to meet the legitimacy condition of [ICCPR] article 19(3) . . . given that [it] protects individuals and their right to freedom of expression and opinion; neither article 19(3) nor article 18 of the Covenant protect ideas or beliefs from ridicule, abuse, criticism or other 'attacks' seen as offensive."); *Addendum to 2011 Special Rapporteur Report to Human*

tion, laws designed to protect governments or their officials from criticism do not reflect legitimate aims.[183] Speech infringements that are adopted to promote the "homogenization of society" similarly fail the legitimacy test.[184] In addition, laws that restrict freedom of expression to promote "civility," "public solidarity," "the State or public interest," or to guard against "misinformation" or "distorted truth" also fail the legitimacy test.[185] The government's failure to demonstrate a legitimate objective when restricting speech has repeatedly contributed to HRC decisions that such bans are unlawful.[186]

Rights Council, *supra* note 154, at ¶ 1173 (explaining that Iran's restrictions on speech based on concerns about "enmity against God," "insulting Islamic sanctities," and "distribution of pictures and materials intended to mock sanctities" are not legitimate public interest objectives); *Special Rapporteur Views on the Maldives*, *supra* note 166, at 6 (critiquing a law that criminalizes criticism of Islam because such an aim is "inconsistent with international human rights law"); Organization for Security and Cooperation in Europe, *Joint Declaration on Universality and Freedom of Expression*, at ¶ 1.f (May 6, 2014), https://bit.ly/2TMP27b [https://perma.cc/VWR3-C7B4] [hereinafter *Universality Joint Declaration*] (noting that "[c]ertain types of legal restrictions on freedom of expression can never be justified by reference to local traditions, culture and values . . . . These include [l]aws which protect religions against criticism or prohibit the expression of dissenting religious beliefs").

183. *See, e.g.*, Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. UGA 4/2012, at 3–4 (June 15, 2012), https://bit.ly/3r1REKr [https://perma.cc/2VVQ-583W] (noting the illegitimacy of infringing on expression to restrict "discussion of government policies and political debate; reporting on human rights, government activities and corruption in government; engaging in election campaigns, peaceful demonstrations or political activities, including for peace or democracy; and expression of opinion and dissent"); *Universality Joint Declaration*, *supra* note 182, at ¶ 1.f (highlighting the illegitimacy of laws that have as their aim to "prohibit debate about issues of concern or interest to minorities and other groups which have suffered from historical discrimination . . . [or] provide for special protection against criticism for officials, institutions, historical figures, or national or religious symbols").

184. *Special Rapporteur 2018 Views on China*, *supra* note 174, at 5–6 (noting that where a regulation's "stated aim is to make 'religion more Chinese and under law, and actively guide religions to become compatible with socialist society,'" such a goal of making "religion more Chinese" is not legitimate).

185. *Special Rapporteur 2016 Report to UNGA*, *supra* note 173, at ¶ 27.

186. *See, e.g.*, Zalesskaya v. Belarus, Commc'n No. 1604/2007, U.N. Doc. CCPR/C/101/D/1604/2007, ¶ 10.5 (Hum. Rts. Comm. 2011) (finding that Belarus failed to meet its burden of showing the ICCPR Article 19 tripartite test had been met because, among other things, it had "not contested the author's assertion that the distributed newspapers and leaflets did not contain information that might harm the rights or reputation of others, did not disclose State secrets, and did not contain calls to disrupt public order or to infringe upon public health or morals"); Kovalenko v. Belarus, Commc'n No. 1808/2008, U.N. Doc. CCPR/C/108/D/1808/2008, ¶ 8.6 (Hum. Rts. Comm. 2013) (finding the State Party had failed to demonstrate its sanctions on an author for "publicly expressing his negative attitude to the Stalinist repressions in Soviet Russia" were imposed for any legitimate reason

3. The Necessity Test

In interpreting ICCPR Article 19(3)'s necessity test, the U.N. human rights machinery has made clear that any restrictive measures must be, among other things, "the least intrusive instrument" to achieve the legitimate aim.[187] The U.N. Special Rapporteur has noted that governments "must demonstrate that the restriction imposes the *least* burden on the exercise of the right and actually protects, or is likely to protect, the legitimate State interest at issue."[188] To ensure a speech restriction is the least intrusive means of achieving a legitimate public interest objective under ICCPR Article 19(3), a State Party should engage in a three-part inquiry:

(1) Is there a way to achieve the desired public interest goal without infringing on speech?

(2) If not, what means are available to achieve the goal, and which one produces the least intrusion on speech interests?

(3) Is the selected infringement effective in advancing the public interest goal?[189]

In reviewing these three questions, it becomes clear that if the legitimate objective can be achieved through non-censorial means, then speech regulators must implement those means rather than infringing on speech. If non-censorial means are insufficient, then

---

set forth in ICCPR Article 19(3)); Youbko v. Belarus, Commc'n No. 1903/2009, U.N. Doc. CCPR/C/110/D/1903/2009, ¶ 9.6 (Hum. Rts. Comm. 2014) (noting that the government had "not explained how, in practice, criticism of a general nature regarding the administration of justice would jeopardize the court rulings at issue, for purposes of one of the legitimate aims set out in [ICCPR Article 19(3)]"); Strambrovsky v. Belarus, Commc'n No. 1987/2010, U.N. Doc. CCPR/C/112/D/1987/2010, ¶ 7.6 (Hum. Rts. Comm. 2014) (finding that the government had failed to show how a one-person picket could hinder public security or order); Kozlov v. Belarus, Commc'n No. 1949/2010, U.N. Doc. CCPR/C/113/D/1949/2010, ¶¶ 7.5-7.6 (Hum. Rts. Comm. 2015) (noting that the government had failed to show how a picket would jeopardize any legitimate public interest objective).

187. GC 34, *supra* note 128, at ¶ 34. In addition to the "least intrusive instrument" test, this part of the tripartite test "'also implies an assessment of the proportionality' of those restrictions. A proportionality assessment ensures that restrictions 'target a specific objective and [do] not unduly intrude upon the other rights of targeted persons.'" David Kaye, (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. USA 6/2017, at 3 (May 9, 2017), https://bit.ly/3qYmsvD [https://perma.cc/94XU-FWYH]. In addition, the proportionality analysis means "[t]he ensuing 'interference with third parties' rights must [also] be limited and justified in the light of the interest supported by the intrusion.'" *Id.*

188. *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 7 (emphasis added).

189. Aswad, *supra* note 9, at 47. The U.N. Special Rapporteur endorsed this trilogy of questions as a means of assessing the necessity of a speech restriction. *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 52 (citing favorably to this framework of inquiries to determine if a company meets the necessity test in content moderation).

speech regulators must select the least intrusive of the available options to achieve the public interest goal. In addition, speech regulators need to assess the effectiveness of the measures selected. If the adopted measure is ineffective or even counter-productive, then that measure is not necessary to achieve the legitimate goal. The U.N. human rights machinery has also referred to such a "necessity" analysis as "narrowly tailoring" restrictions.[190]

In assessing if speech restrictions are necessary, the U.N. human rights machinery has used this trilogy of inquiries to determine whether a speech restriction meets the least intrusive means test. For example, in the context of disinformation, the Special Rapporteur critiqued a Qatari law that punished the publication of false news not only on vagueness grounds but also because non-censorial methods existed for combatting falsehoods, such as "the promotion of independent fact checking mechanisms . . . and public education and media literacy."[191] Similarly, in a study involving online disinformation during elections, the Special Rapporteur advocated for consideration of non-censorial methods that promote an enabling environment for freedom of expression, such as "heightened transparency regarding advertisement placement and sponsored content," independent fact checking, and media literacy initiatives.[192] A joint declaration of the Special Rapporteur and regional free expression watchdogs likewise highlighted that platforms should seek to deploy non-censorial tools to combat disinformation, including developing advertising archives, providing public alerts, and engaging in greater transparency with respect to platforms' use of automated tools to curate content.[193]

If non-censorial methods are insufficient to achieve the public purpose, the U.N. machinery has called upon speech regulators to consider a range of available options and select the least intrusive means to achieve the objective, which involves an examination of the particular context at stake. The Special Rapporteur has frequently stated that restrictions on speech that may chill expression should not be deployed "when less invasive techniques are availa-

190. *See Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 47 (observing that "[g]ranular data on actions taken will also establish a basis to evaluate the extent to which companies are narrowly tailoring restrictions").

191. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Comm'n No. OL QAT 1/2020, at 3 (Apr. 14, 2020), https://bit.ly/2VfrCaM [https://perma.cc/P88R-NCNG].

192. *Elections in the Digital Age*, *supra* note 161, at 11.

193. *Joint Elections Declaration*, *supra* note 161, at 4.

ble or have not yet been exhausted."[194] For example, the Special Rapporteur has found that states have failed to meet their burden of showing the necessity of "backdoors" that weaken or bypass encryption "given a wide range of investigative tools at their disposal."[195] Similarly, in providing guidance to social media platforms, the Special Rapporteur has highlighted that such companies have a wide range of options from which to select when dealing with harmful content and that they should bear the burden of demonstrating publicly that they have selected the least intrusive option.[196]

The U.N. human rights machinery has also called on speech regulators to engage in evidence-based assessments of whether selected restrictions are effective in achieving legitimate public interest objectives, as ineffective or counter-productive restrictions are not justifiable infringements on speech.[197] For example, in the context of a study on false information and elections, the Special Rapporteur has noted "approaches for combatting disinformation should be evidence-based and tailored to proven or documented impacts of disinformation or propaganda."[198] In the context of ex-

194. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 83, U.N. Doc. A/HRC/23/40 (Apr. 17, 2013); *see also id.* ¶ 50 (noting that least intrusive measures are often not selected in the context of communications surveillance programs).

195. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Research Paper 1/2018: Encryption and Anonymity Follow-Up Report*, ¶ 13 (June 2018), https://bit.ly/2TyyNKP [https://perma.cc/Z5RQ-QPAJ].

196. *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 51. The Special Rapporteur noted that there are a broad range of tools that platforms can deploy to tackle harmful speech:

> They can delete content, restrict its virality, label its origin, suspend the relevant user, suspend the organization sponsoring the content, develop ratings to highlight a person's use of prohibited content, temporarily restrict content while a team is conducting a review, preclude users from monetizing their content, create friction in the sharing of content, affix warnings and labels to content, provide individuals with greater capacity to block other users, minimize the amplification of the content, interfere with bots and coordinated online mob behavior, adopt geolocated restrictions and even promote counter-messaging.

Id.; *see also* Aswad, *supra* note 9, at 49 (describing a continuum of options available to social media companies in selecting the least intrusive means when curating speech on their platforms).

197. As noted by the Special Rapporteur, "States must demonstrate that the [speech] restriction imposes the least burden on the exercise of the right and actually protects, or is likely to protect, the legitimate interest at issue. States may not merely assert necessity but must demonstrate it." *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 7.

198. *Elections in the Digital Age*, *supra* note 161, at 11.

amining "encroachments on encryption and anonymity," the Special Rapporteur stated that "'a detailed and evidence-based public justification' is critical."[199] In an amicus brief before the Korean Constitutional Court regarding governmental access to user data, the Special Rapporteur similarly observed that assertions of the necessity to access user data must be substantiated by "'a detailed and evidence-based public justification.'"[200] When expressing concerns to the Brazilian government about a bill that would prohibit "political and ideological indoctrination," the Special Rapporteur took the position that the bill failed the necessity test as there was no "empirical evidence or findings indicating a need for these bills . . . there appears to be no reason to believe that other educational practices" could not serve as less intrusive ways forward.[201]

The U.N. human rights machinery has also identified a variety of areas in which restrictions or particular sanctions on speech will always or almost always fail the necessity test. For example, the U.N. human rights machinery has consistently called for the decriminalization of defamatory or otherwise false speech.[202] The U.N. Special Rapporteur has also called for the abolition of "prior-

199. *2015 Special Rapporteur Report to Human Rights Council*, *supra* note 173, at ¶ 35.

200. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Third-Party Intervention Filed in the Constitutional Court of the Republic of Korea*, ¶¶ 24–25, (May 9, 2017), https://bit.ly/3jPBlP6 [https://perma.cc/6JZ4-N6CK].

201. *Joint Special Rapporteur Views on Brazil*, *supra* note 167, at 4–5, 7.

202. *See, e.g.*, GC 34, *supra* note 128, at ¶ 47 (recommending the decriminalization of defamation laws); *Special Rapporteur Views on Israel and the Occupied Palestinian Territory*, *supra* note 154, at ¶ 53 ("The Special Rapporteur has consistently called for decriminalization of defamation as a criminal offence, which is inherently harsh and encourages self-censorship."); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on His Visit to Ethiopia*, ¶ 66, A/HRC/44/49/Add.1 (Apr. 29, 2020) [hereinafter *Special Rapporteur Views on Ethiopia*] (noting the Special Rapporteur "believes that the use of criminal sanctions is generally inappropriate to address false news, and that imprisonment is never an appropriate penalty. He urges the authorities to decriminalize the offence of defamation and to provide for reasonable civil liabilities"); *Affidavit for Third-Party Intervention Filed Before ECOWAS Court*, *supra* note 152, at 14 (commemorating the Special Rapporteur's view that "[l]ibel, defamation, and 'false news' laws generally fail to be proportionate when they (i) carry criminal punishments, (ii) do not provide wider latitude for criticism against public officials . . . and/or (iii) do not provide wider latitude for speech in the public interest . . ."); Lydia Cacho Ribiero v. Mexico, Commc'n No. 2767/2016, U.N. Doc. CCPR/C/123/D/2767/2016, ¶ 10.8 (Hum. Rts. Comm. Aug. 29, 2018) (observing that defamation should not be criminalized and explaining that "[i]f defamation should never result in a penalty of deprivation of liberty being imposed . . . then a fortiori no detention based on charges of defamation may ever be considered either necessary or proportionate").

censorship bodies,"[203] highlighted that a variety of prior restraints on expression do not meet the necessity test,[204] and emphasized that governmental regulation of (or pressure on) platforms to develop filters that prevent individuals from uploading content would

203. Farida Shaheed (Special Rapporteur in the Field of Cultural Rights) et al., Special Rapporteurs' Joint Commc'n No. EGY 9/2015, at 9–10 (Aug. 19, 2015), https://bit.ly/3xmMtqR [https://perma.cc/JK6Q-L32Z] (calling for the abolition of "prior-censorship bodies or systems," explaining that "[p]rior censorship should be a highly exceptional measure, undertaken only to prevent the imminent threat of grave irreparable harm to human life or property," and noting that post-publication liability should "be imposed exclusively by a court of law"); *Special Rapporteur Views on Israel and the Occupied Palestinian Territory*, *supra* note 154, at ¶ 25 (expressing concern at the existence of a prior censorship body, noting no country should have such a body, and reminding that limitations on free expression must, among other things, be "the least restrictive means available to protect a specific and legitimate national security interest"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 50, U.N. Doc. A/69/335 (Aug. 21, 2014), ("The imposition of prior censorship to protect children from harmful material provides an example of disproportionate restrictions that run counter to international human rights standards."); *Special Rapporteur Views on the Maldives*, *supra* note 166, at 7 (condemning a prior censorship regime in which "all literary works, including poetry, to be submitted to the Government for approval prior to publication, and failure to comply can result in fines"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. KEN 10/2017, at 5 (July 26, 2017), https://bit.ly/2UtRsrf [https://perma.cc/7RFA-FP3U] (criticizing a regime that entailed "a minimum of 48 hours for flagging messages for review" as creating "in effect prior restraint" on freedom of expression).

204. *Special Rapporteur Report on Tajikistan*, *supra* note 158, at ¶ 32 (finding that "blanket shutdowns of entire social media sites are neither necessary nor proportionate to protect public order or national security"); *Special Rapporteur Views on Ethiopia*, *supra* note 202, at ¶ 51 (explaining that Internet shutdowns fail the necessity test because "they affect areas beyond the Government's specific concerns and cut users off from a variety of essential activities and services such as emergency services and health information, mobile banking and commerce, transportation, school classes, voting and election monitoring, reporting on major crises and events, and human rights investigations"); *Joint Declaration on Expression and the Internet*, *supra* note 178, at ¶ 3 ("Mandatory blocking of entire websites, IP addresses, ports, network protocols or types of uses (such as social networking) is an extreme measure—analogous to banning a newspaper or broadcaster—which can only be justified in accordance with international standards . . . ."); *Special Rapporteur Views on China*, *supra* note 157, at 5 (criticizing a draft law that would give the government broad discretion to "shut down internet communication" in the face of threats to national security, public order, or other "major security interests"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. CAN 1/2018, at 5 (Apr. 17, 2018), https://bit.ly/3dMReSG [https://perma.cc/4338-NAR6] ("While the enforcement of copyright law may be a legitimate aim, I am concerned that website/application blocking is almost always a disproportionate means of achieving this aim.").

violate the necessity test.[205] A variety of bans on criticism of the government have also not met the necessity test.[206]

In sum, U.N. monitoring mechanisms have rigorously interpreted and enforced ICCPR Article 19's tripartite test of legality, legitimacy, and necessity, forcing speech regulators to bear the burden of demonstrating the validity of restrictions on expression from various angles. This Section now turns to the intersection of U.N. free speech standards and hate speech.

C. Hate Speech

The U.N. human rights system contains *mandatory* bans on certain forms of hate speech and allows restrictions on other types of hate speech (i.e., *discretionary* hate speech bans). However, under either scenario, governments must demonstrate that the bans pass ICCPR Article 19(3)'s tripartite test. Part II(C) begins by analyzing the scope of the mandatory hate speech bans in the ICCPR and then turns to the Convention on the Elimination of Racial Discrimination ("CERD"). Next, this Part assesses how the application

---

205. *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 34 (highlighting that governmental efforts towards automated tools to filter content before it is uploaded to the Internet "would serve as a form of pre-publication censorship"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. OTH 41/2018, at 7 (June 13, 2018), https://bit.ly/3dOTXLD [https://perma.cc/5SR3-AGYS] (criticizing a proposed EU directive as "pre-publication censorship" as it would incentivize platforms "to restrict at the point of upload user-generated content that is perfectly legitimate and lawful"); *Joint Declaration on Expression and the Internet*, *supra* note 178, at ¶ 3.b ("Content filtering systems which are imposed by a government or commercial service provider and which are not end-user controlled are a form of prior censorship and are not justifiable as a restriction on freedom of expression."); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. PAK 3/2020, at 5 (Mar. 19, 2020), https://bit.ly/3yuVbUc [https://perma.cc/K9RF-A9CB] (criticizing a regime that "would potentially legitimize the prior censorship of content by social media companies" by requiring prevention of live-streaming of prohibited content).

206. *See, e.g.*, *Special Rapporteur Report on Turkey*, *supra* note 158, at ¶ 85 ("The criminalization of individuals solely for criticism of the Government can never be considered a necessary restriction to freedom of expression."); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Expression), Special Rapporteur Commc'n No. RUS 2/2018, at 3 (Feb. 5, 2018), https://bit.ly/36qXzzl [https://perma.cc/PPU3-D3TF] ("The penalization of a media outlet, publishers or journalists solely for being critical of the government can never be considered a necessary restriction on freedom of expression . . . ."); Yashar Agazade and Rasul Jafarov, Commc'n No. 2205/2012, U.N. Doc. CCPR/C/118/D/2205/2012, ¶ 7.4 (Hum. Rts. Comm. Mar. 16, 2017) ("The penalization of a media outlet or journalist solely for being critical of the government or the political social system espoused by the Government can never be considered a necessary restriction of freedom of expression.").

of ICCPR Article 19(3)'s tripartite test impacts the scope of both mandatory and discretionary prohibitions on hate speech. The Section concludes that (1) the U.N. machinery's interpretations of U.N. treaty provisions concerning hate speech from the last decade have significantly narrowed the potential breadth of mandatory hate speech bans and (2) the strict application of ICCPR Article 19(3)'s tripartite test to all speech restrictions has further confined the reach of both mandatory and discretionary hate speech bans.

1. Mandatory Hate Speech Bans

ICCPR Article 20 provides that "[a]ny *advocacy* of national, racial or religious hatred that constitutes *incitement* to *discrimination, hostility or violence* shall be prohibited by law."[207] Though the HRC has not issued a General Comment on the scope of this article, the Special Rapporteur has recommended a number of relevant interpretations. First, he noted that, by its own terms, the article requires the prohibition (i.e., civil sanctions), but does not require the criminalization of speech.[208] Second, speech must meet 3 conditions to qualify as proscribed under Article 20: (1) there must be advocacy of national, racial, or religious hatred (2) that rises to the level of incitement to (3) the three harms of discrimination, hostility, or violence.[209]

The Special Rapporteur has advised that the thresholds for each of these conditions should be high and has highlighted appropriate definitions of key terms. For example, with respect to "advocacy," the speaker should have an *intent* to promote hatred towards a particular group and engage in such advocacy publicly.[210] Incitement should be understood to require an "*imminent*" and *likely* risk of harm to the targeted group.[211] Regarding the harms, the Special

207. ICCPR, *supra* note 115, at art. 20 (emphasis added). When the United States joined the ICCPR, it took a reservation to Article 20, which states this article "does not authorize or require legislation or other action by the United States that would restrict the right of free speech and association" under U.S. law. ICCPR Treaty Collection, *supra* note 120. The phrasing is notable as it does not explicitly commemorate that Article 20 goes beyond the First Amendment, but by virtue of being styled a "reservation," it does indicate concern that Article 20 could be interpreted in ways inconsistent with U.S. law.

208. *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 47.

209. *Id.* at ¶ 43

210. *Id.* at ¶ 44(b).

211. *Id.* at ¶¶ 44(c), 45(e) (emphasis added); *see also* David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. POL 2/2018, at 4 (Feb. 13, 2018), https://bit.ly/3xzfWxS [https://perma.cc/JK3L-B26Z] (observing that to the extent a Polish law was seeking to counter national hatred under ICCPR Article 20, the law inappropriately neglected the incitement standard, which requires as-

Rapporteur noted the ambiguity of "hostility" and recommended it be viewed as "a *manifestation* of hatred *beyond* a mere state of mind."[212]

In addition to the Special Rapporteur's interpretations of Article 20, the U.N. Office of the High Commissioner for Human Rights released a report of an independent experts group that it had convened to provide recommendations on this article.[213] This document, known as the Rabat Plan of Action, emphasized non-censorial means of tackling hate speech[214] and also set forth a framework of factors to assess when the gravity of the speech could render criminal bans "necessary."[215] Those factors are (1) the social and political context when the speech was made; (2) the status of the speaker; (3) the intent of the speaker (noting negligence and recklessness would not suffice); (4) the content and form of the speech; (5) the reach of the speech; and (6) the likelihood of harm, including its imminence.[216] In 2019, the Special Rapporteur recommended use of the Rabat factors in assessing speech that is most likely to cause harm and thus most eligible for restriction.[217]

The CERD, a treaty with 182 State Parties that predates the ICCPR and seeks to end racial discrimination, contains an additional mandatory ban on hate speech.[218] CERD Article 4 provides that states, with "due regard" to other human rights such as freedom of expression, must:

> declare an offense punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all . . . incitement to [violence] against any race or group of persons of another colour or ethnic origin, and also

---

sessment of various factors, including "the intent of the speaker, the form, style, and magnitude of the expression, and the likelihood of harm occurring (including its imminence)").

212. *Special Rapporteur 2012 UNGA Report, supra* note 171 at ¶ 44(e) (emphasis added).

213. *The Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility, or Violence*, A/HRC/22/17/Add.4, ¶ 1 (Oct. 5, 2012).

214. *See id.* at ¶¶ 35–41 ("[L]egislation is only part of the larger toolbox to respond to the challenges of hate speech.").

215. *Id.* at ¶ 29.

216. *Id.*

217. *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 57.

218. *International Convention on the Elimination of Racial Discrimination*, UNITED NATIONS TREATY COLLECTION, https://bit.ly/3dQPVCh [https://perma.cc/FH22-7XY2] (last visited Aug. 11, 2021) [hereinafter CERD Treaty Collection].

> the provision of any assistance to racist activities, including the financing thereof.[219]

Though the CERD does not cover religious groups, its Article 4 is broader than ICCPR Article 20 because CERD Article 4 requires the criminalization of speech and appears to mandate bans on "disseminating" certain racially intolerant speech, even if such expression is untethered to incitement.

However, after the issuance of the HRC's General Comment 34, the U.N. Committee on the Elimination of Racial Discrimination ("the CERD Committee"), the expert body charged with monitoring implementation of the CERD, advised in 2013 that both types of Article 4 offenses (e.g., *dissemination* of speech based on racial superiority and *incitement* to racial violence/discrimination) require an assessment of the intention of the speaker and the likelihood of imminent harm.[220] In addition to holding dissemination offenses to an incitement standard, the CERD Committee further emphasized that the criminalization of racist hate speech "should be reserved for serious cases" and should not cover ideas in academic or political debates that do not rise to the level of incitement.[221] By interpreting both dissemination and incitement offenses to require an incitement analysis (i.e., (1) the speaker's intention to create harm and an evaluation of (2) the likelihood of (3) imminent harm), the CERD Committee's 2013 interpretations rep-

---

219. CERD, *supra* note 122, at art. 4(a) (emphasis added). The CERD also provides that state parties are to "declare illegal and prohibit organizations, and also organized and all other propaganda activities, which promote and incite racial discrimination, and shall recognize participation in such organizations or activities as an offence punishable by law." *Id.* at art. 4(b). When it joined the CERD, the United States took a reservation to Article 4 that stated it did not accept any obligation contrary to U.S. law. CERD Treaty Collection, *supra* note 218.

220. Comm. on the Elimination of Racial Discrimination, General Recommendation No. 35, ¶ 16, U.N. Doc. CERD/C/GC/35 (Sept. 26, 2013) [hereinafter GR 35] (recommending that, when determining if dissemination and incitement offenses properly fall within CERD Article 4, "the intention of the speaker, and the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech in question" should be taken into account). Former CERD Committee Member (2001–2014) and lead drafter of CERD GR 35, Patrick Thornberry observes that this General Recommendation "decisively rejects any suggestion of a 'strict liability' approach to dissemination and incitement . . . [by linking] them with principles of criminal law on mental elements in crime." Patrick Thornberry, *International Convention on the Elimination of All Forms of Racial Discrimination: The Prohibition of 'Racist Hate Speech*,' *in* THE UNITED NATIONS AND FREEDOM OF EXPRESSION AND INFORMATION: CRITICAL PERSPECTIVES 121, 131 (Cambridge University Press 2015). Thornberry notes that the inclusion of *imminence* also narrows "the scope of potential hate speech prosecutions." *Id.* at 132.

221. GR 35, *supra* note 220, at ¶¶ 12, 25.

resented a significant and purposeful narrowing of its approach to Article 4.[222]

In addition, the U.N. human rights machinery has taken the position that all bans on speech—including the mandatory bans in ICCPR Article 20 and CERD Article 4—are subject to ICCPR Article 19's tripartite test.[223] For example, in its 2011 General Comment, the HRC stated that any restriction on speech—whether imposed under ICCPR Article 20 or otherwise—must meet Article 19's tripartite test.[224] Similarly, the CERD Committee has stated

---

222. *See* PATRICK THORNBERRY, THE INTERNATIONAL CONVENTION ON THE ELIMINATION OF ALL FORMS OF RACIAL DISCRIMINATION 297–98 (Oxford 2016). Thornberry concludes that

> the fresh reading of Article 4 takes the Convention closer to the ICCPR . . . . Overall, it may be argued that [General Recommendation] 35 takes CERD practice nearer to 'libertarian' currents regarding the prosecution of hate speech crimes; the suggested criminal law requirement of the need for 'imminence' of the consequences of incitement may be more stringent than in many jurisdictions.

Id. at 301–02. Thornberry notes that the CERD Committee's prior views, which did not apply the thresholds in General Recommendation 35, are no longer valid. *Id.* at 293–94.

223. *See supra* notes 151–206 and accompanying text.

224. GC 34, *supra* note 128, at ¶¶ 50-52 ("[A] limitation that is justified on the basis of article 20 must also comply with article 19, paragraph 3 . . . . In *every* case in which the State restricts freedom of expression it is necessary to justify the prohibitions . . . in strict conformity with article 19.") (emphasis added). A review of the Human Rights Committee's recommendations to ICCPR State Parties about their implementation of treaty obligations reveals that the Committee has sometimes called for restricting hate speech in accordance with Articles 19 and 20 but at other times has called for restricting hate speech without specifically mentioning those articles. *See, e.g.*, U.N. Hum. Rts. Comm., *Concluding Observations on the Fourth Periodic Report of Czechia*, ¶ 17 U.N. Doc. CCPR/C/CZE/CO/4 (Dec. 6, 2019) (calling for the prohibition of hate speech in accordance with ICCPR Articles 19 and 20); U.N. Hum. Rts. Comm., *Concluding Observations on the Fourth Periodic Report of Algeria*, ¶ 20 U.N. Doc. CCPR/C/DZA/CO/4 (Aug. 17, 2018) (calling on Algeria "to combat hate speech by public or private persons, including on social media and the Internet, in accordance with articles 19 and 20 of the [ICCPR] and general comment No. 34"); U.N. Hum. Rts. Comm., *Concluding Observations on the Sixth Periodic Report of Hungary*, ¶ 18 U.N. Doc. CCPR/C/HUN/6 (May 9, 2018) (calling for the prohibition of "any advocacy of ethnic or racial hatred that constitutes incitement to discrimination, hostility, or violence" without mention of ICCPR Article 19 safeguards); U.N. Hum. Rts. Comm., *Concluding Observations on the Fourth Periodic Report of Slovakia*, ¶ 15 U.N. Doc. CCPR/C/SVK/CO/6 (Nov. 22, 2016) (calling on Slovakia to "adopt measures to tackle hate speech on the grounds of sexual orientation and gender identity" without mentioning ICCPR Articles 19 and 20 or General Comment 34). In other instances, the Committee has called for "taking measures" against hate speech without explaining whether such measures should involve restricting speech or focus on good governance measures to promote tolerance (e.g., inter-faith dialogues). *See, e.g.*, U.N. Hum. Rts. Comm., *Concluding Observations on Nigeria in the Absence of Its Second Periodic Report*, ¶ 45 U.N. Doc. CCPR/C/NGA/CO/2 (Aug. 29, 2019) (noting Nigeria "should take measures against discrimination and

that restrictions on racist hate speech must also comply with ICCPR Article 19's legality and necessity tests.[225] Over the last decade, the U.N. Special Rapporteur has also emphasized that restrictions pursuant to ICCPR Article 20 or CERD Article 4 must pass the ICCPR Article 19(3) tripartite test.[226] This Section now turns to whether and how the U.N. human rights machinery's application of the tripartite test has tempered the potential scope of both mandatory and discretionary bans on hateful speech.

2. The Impact of the ICCPR Article 19 Tripartite Test

In applying the legality prong of the tripartite test, the U.N. system has consistently condemned as vague a variety of bans on hateful speech, including those that contain similar phrasing to the text of ICCPR Article 20(2) and CERD Article 4. For example, U.N. experts have found that laws banning the *incitement* of hatred, discrimination, unrest, fear, or hostility to constitute inappropriately vague bans.[227] In addition, the U.N. machinery has found that

---

hate speech and incitement to hatred and violence aimed at any religious community" without specifying the nature of such measures). Despite the Committee's use of abbreviated phrasing in its concluding observations, its recommendations should be understood and implemented within the context of and harmoniously with existing interpretations of the UN's human rights machinery (such as General Comment 34) rather than reflecting an abdication or repudiation of such interpretations.

225. *See* GR 35, *supra* note 220, at ¶¶ 12, 19 ("The application of criminal sanctions should be governed by principles of legality, proportionality and necessity."). Given the public interest objective for restricting speech is set forth in the CERD, the Committee likely decided not to mention the "legitimacy" test. *See also Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 15 (noting that the CERD Committee clarified in 2013 that "the 'due regard' language in article 4 of the Convention as meaning that strict compliance with freedom of expression guarantees is required").

226. *See, e.g.*, *Special Rapporteur 2019 Report*, *supra* note 140, at ¶¶ 13, 16 (stressing the applicability of ICCPR Article 19's tripartite test to all speech bans, including CERD Article 4 and ICCPR Art. 20); *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 41 (emphasizing the applicability of ICCPR Article 19 to all mandatory hate speech bans).

227. *See, e.g.*, *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 51 (criticizing as a vague offense "'inciting violence against a religious authority' in Angola, 'causing national, racial, or religious hate, discord and intolerance' in . . . Macedonia . . . and 'misrepresenting events and inciting violence' in Somalia"); *Joint Declaration on Media Independence*, *supra* note 165, at ¶ 3.f (calling for bans on "incitement to hatred" to be "defined clearly and narrowly"); *Special Rapporteur 2016 Report to UNGA*, *supra* note 173, at ¶ 25 (criticizing a Pakistani law that penalizes information "that advances or is likely to advance inter-faith, sectarian or racial hatred" as vague and also observing that regional "European human rights law also fails to define hate speech adequately"); *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 32 (criticizing Germany's Network Enforcement Act for requiring platforms to enforce vague provisions in its penal code such as speech bans on persons who "in a manner *capable* of disturbing the public peace, *incites*

laws prohibiting the *spread* of hateful, hostile, divisive, and/or racist views to fail the legality test for vagueness.[228] U.N. human rights mechanisms have also found that laws banning the *creation* of animosity, hatred, or discord[229] as well as laws prohibiting disrespect for human dignity[230] to constitute inappropriately vague laws in

hatred against a national, racial, religious group or a group defined by their ethnic origins"); *Special Rapporteur Views on Jordan*, *supra* note 157, at 2 (finding that a cybersecurity law that defined hate speech as "any statement or act that would incite discord (*fitna*), religious, sectarian, racial or ethnic strife or discrimination between individuals or groups" was unduly vague and overbroad); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. EGY 1/2014, at 5 (Jan. 14, 2014) https://bit.ly/3jQK3wO [https://perma.cc/4ASR-FBSM] (expressing concerns about the "unclear wording of exceptions relating to the crimes of 'incitement of violence, discrimination between citizens"); *2011 Special Rapporteur Report to UNGA*, *supra* note 155, at ¶¶ 26, 29 (noting vagueness concerns with hate speech formulations such as "incitement to religious unrest," "promoting divisions between religious believers and non-believers," and inciting "offenses that damage public tranquility"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Joint Special Rapporteur Commc'n No. VEN 9/2017, at 4 (Jan. 11, 2018), https://bit.ly/36iDL0Y [https://perma.cc/T4TM-8NS3] (translated text) (criticizing as unduly vague Venezuelan bans on the promotion or incitement to hatred or discrimination).

228. *See, e.g.*, David Kaye (Special Rapporteur on the Right of Freedom of Expression and Opinion), Special Rapporteur Commc'n No. EGY 13/2018, at 1–2 (Aug. 9, 2018), https://bit.ly/3jQK7N4 [https://perma.cc/J7RS-4PJP] (criticizing as vague an Egyptian law that authorizes suspension of a website, blog, or social media account if, among other things, it "spreads hateful views"); *Special Rapporteur 2016 Report to UNGA*, *supra* note 173, at ¶ 25 (criticizing "vague prohibitions on 'advocacy of hatred' that do not amount to incitement under [ICCPR] Article 20 . . . or meet the requirement of necessity under [ICCPR] Article 19(3)"); *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 52 (criticizing as unduly vague bans on "expression of feelings of hostility" and "exciting racial hostility"); David Kaye (Special Rapporteur on the Right of Freedom of Expression and Opinion), Special Rapporteur Commc'n No. MRT 5/2017, at 8 (Jan. 24, 2018), https://bit.ly/3xxQVTC [https://perma.cc/W94C-GUAT] (translated text) (criticizing as vague Mauritania's ban on speech that (1) has a "racist nature," (2) "supports or communicates terms that could reveal an intention to leave or incite to hurt morally or physically, promote or incite to hatred," or (3) "incites discrimination, hatred or violence, defamation and insult on the grounds of origin or belonging racial, ethnic, nationality").

229. *See, e.g.*, *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 51 (noting the vagueness of Macedonia's ban on "causing national, racial or religious hate, discord and intolerance"); *2018 Special Rapporteur Comment on BGD*, *supra* note 156, at 4 (noting concerns about draft bill's vague bans on speech that "creates animosity, hatred or antipathy among the various classes and communities"); *2018 Special Rapporteur Comment on Malaysia*, *supra* note 154, at 3 (criticizing as vague bans on any speech that "conjures feelings of 'hatred,' 'contempt,' 'disaffection,' 'discontent,' 'ill will,' or 'hostility'").

230. *See* Addendum to *2011 Special Rapporteur Report to Human Rights Council*, *supra* note 154, at ¶ 845 (criticizing as overly broad a Hungarian requirement that media produce content that respects "human dignity" and prohibits content that is "self-gratifying" as well as "detrimental coverage of persons in

contravention of ICCPR Article 19(3). The Special Rapporteur has similarly observed that social media "policies on hate, harassment and abuse also do not clearly indicate what constitutes an offence."[231] The U.N. machinery has expressed concern that vague hate speech bans "are in fact used to suppress critical and opposing voices"[232] as well as contain an unacceptably high risk of misuse.[233]

In applying ICCPR Article 19(3)'s necessity test to particular hate speech laws, the U.N. human rights machinery has frequently applied the trilogy of questions discussed above in assessing whether a speech ban constitutes the "least intrusive means" of achieving a legitimate objective.[234] For example, U.N. experts have pressed governments to address hatred and tolerance through non-censorial means, such as education, counter-speech, outreach by public officials, and enforcement of discrimination and hate crimes

---

humiliating or defenseless situations"); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), Special Rapporteur Commc'n No. IND 3/2019, at 3 (Feb. 14, 2019), https://bit.ly/3k8fvGH [https://perma.cc/3U4C-M3H6] (criticizing as vague a proposed rule in India that would ban "information that is 'racially, ethnically objectionable, disparaging'"); *Joint Communication on Spanish Legislation*, *supra* note 158, at 5–6 (finding a ban on expression that includes "discredit, contempt or humiliation of victims of terrorist crimes or their relatives" to be vague and overly broad).

231. *Special Rapporteur 2018 Report*, *supra* note 10, at ¶ 26 ("Twitter's prohibition of "behavior that incites fear about a protected group and Facebook's distinction between 'direct attacks' on protected characteristics and merely 'distasteful or offensive content' are subjective and unstable bases for content moderation."); *see also Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 46 (criticizing as vague the messaging app WeChat's ban on "content . . . which in fact or in our reasonable opinion . . . is hateful, harassing, abusive, racially or ethnically offensive, defamatory, humiliating to other people (publicly or otherwise), threatening, profane or otherwise objectionable"); *Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 46 (criticizing as vague (1) a Russian social network's speech code that bans speech that "propagandizes and/or contributes to racial, religious, ethnic hatred or hostility, propagandizes fascism or racial superiority" or "contains extremist materials" and (2) American social media companies for speech codes that restrict content against protected groups, but do not define key words such as "promote," "incitement," and "targeting groups").

232. *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 51 (noting misuse of vague hate speech laws in Macedonia to "suppress any criticism of the Macedonian Orthodox Church" and in Somalia "to arrest and detain independent journalists").

233. *Special Rapporteur Views on Ethiopia*, *supra* note 202, ¶¶ 33–34 (noting Ethiopia's Hate Speech and Disinformation Prevention and Suppression Proclamation contained a vague hate speech ban, which prohibited "speech that deliberately promotes hatred, discrimination or attack against a person or a discernable group of identity, based on ethnicity, religion, race, gender or disability," creating "a serious risk that the law may be used to silence critics").

234. *See supra* note 189 and accompanying text (describing the trilogy of questions as whether a state has non-censorial means to address a public interest objective, whether a state has selected the least intrusive measure that burdens speech, and whether a state can demonstrate the intrusion on speech is effective).

laws.[235] It would indeed be peculiar if human rights law rewarded governments that do not engage in such good governance measures by granting them censorial powers as a first resort for tackling hate and intolerance in their societies.

With regard to whether a hate speech ban represents the least intrusive means, the Special Rapporteur has considered various contextual factors, including the following: (1) the likelihood of (2) very near term (i.e., imminent) harm and (3) the speaker's intent.[236]

---

235. Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on His Visit to Italy*, ¶ 64, A/HRC/26/30/Add. 3 (Apr. 29, 2014) (recommending non-legal measures to combat hate such as "that political leaders actively promote tolerance and understanding towards others and support open debates and exchanges of ideas in which everyone can participate on an equal footing [and that public officials] systematically denounce and condemn hate speech publicly"); Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on His Visit to Macedonia*, ¶ 78, A/HRC/26/30/Add.2 (Apr. 1, 2014) (encouraging Macedonia to tackle hate by "investing in non-legal measures, such as education and counter-speech, to encourage the abandonment of discriminatory stereotypes" and stating that "[f]ormal and open rejection of hate speech by high level public officials, in particular hate messages targeting sexual minorities, would also play an important role in the struggle against tolerance and discrimination"); *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶¶ 62–64 (calling on governments to "proactively facilitate counter-speech of individuals belonging to groups that are systematically targeted by hate speech" as well as emphasizing the need for "formal rejections of hate speech by high-level public officials and initiatives to engage in interreligious or intercultural dialogue"); *Special Rapporteur Views on Ethiopia*, *supra* note 202, at ¶ 32 (encouraging Ethiopia to tackle hatred through "regular public messages from high-level officials and community leaders about the danger of hate speech, media literacy, professional training and self-regulation").

236. *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 79. Specifically, the Special Rapporteur has stated:

> To prevent any abusive use of hate speech laws, the Special Rapporteur recommends that only serious and extreme instances of incitement to hatred be prohibited as criminal offences. The Special Rapporteur thus calls upon States to establish high and robust thresholds, *including* the following elements: severity, intent, content, extent, likelihood or probability of harm occurring, imminence and context.

Id. The Special Rapporteur has emphasized near-term harm with respect to the necessity test in other contexts as well. *See, e.g.*, Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶¶ 52–53 U.N. Doc. A/68/362 (Sept. 4, 2013).

> For a restriction to be necessary, it must . . . not be more restrictive than is required for the achievement of the desired purpose or protected right . . . . [T]he authorities must demonstrate, in specific and individualized fashion, the precise nature of the *imminent* threat, as well as the necessity for and the proportionality of the specific action taken. A direct and im-

Specifically, the Special Rapporteur has stated nobody "should be penalized for the dissemination of hate speech unless it has been shown that they did so with the intention of inciting discrimination, hostility or violence."[237] He has emphasized that incitement requires a "real and imminent danger" of particular harm and the intent of the speaker to incite harm.[238] He has reiterated that freedom of expression includes speech that is "offensive, disturbing and shocking" and that "not all types of inflammatory, hateful or offensive speech amount[s] to incitement."[239] In sum, it is challenging for governments to meet their burden of demonstrating hate speech bans are the least intrusive means unless there is a real risk of imminent harm directly related to the speech.

---

mediate connection between the expression (or the information to be disclosed) and the alleged threat must be established.

Id.; Frank LaRue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 36, U.N. Doc. A/HRC/17/27(May 16, 2011) (noting that expression should only be limited for national security or counter-terrorism purposes when "(a) the expression is intended to incite imminent violence, (b) it is likely to incite such violence; and (c) there is a direct and immediate connection between the expression and the likelihood or occurrence of such violence").

237. *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 50.

238. *Id.* at ¶ 46. At times, when issuing recommendations to ICCPR State Parties about their implementation of the treaty, the Human Rights Committee has appeared to depart from requiring such high thresholds. For example, the Committee recommended that Estonia prohibit "the public denial, justification or condoning of genocide, crimes against humanity, war crimes or hate propaganda that is racist or otherwise incites discrimination" without mentioning the high thresholds of intent and imminent danger. U.N. Hum. Rts. Comm., *Concluding Observations on the Fourth Periodic Report of Estonia*, ¶ 14, U.N. Doc. CCPR/C/EST/CO/4 (Apr. 18, 2019). Similarly, the Committee recommended the banning of hateful organizations in Bosnia and Poland without mentioning whether such prohibitions would constitute the least intrusive means and whether the thresholds of intent and imminent danger were met. *See* U.N. Hum. Rts. Comm., *Concluding Observations on the Seventh Periodic Report of Poland*, ¶ 16, U.N. Doc. CCPR/C/POL/CO/7 (Nov. 23, 2016); U.N. Hum. Rts. Comm., *Concluding Observations of the Fourth Periodic Report of Bosnia and Herzegovina*, ¶ 20, U.N. Doc. CCPR/C/BIH/CO/3 (Apr. 18, 2019). As noted previously (*see supra* note 224), the author posits that the Committee's use of abbreviated recommendations when issuing concluding observations to State Parties should be understood and implemented within the context of the U.N. human rights machinery's overall interpretations, including General Comment 34.

239. *Special Rapporteur 2012 UNGA Report*, *supra* note 171, at ¶ 49. *See also Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 48 (noting that "the use of a derogatory term to refer to a national or racial or religious group . . . on its own[ ] would not be subject to restriction under human rights law").

3. Protections Against Discrimination

In addition to the ICCPR Article 19(3) tripartite test, the HRC has made clear that any restriction on freedom of expression must also comport with the ICCPR's other human rights protections in the ICCPR, including those on the right to be free from discrimination (Articles 2 and 26).[240] These articles provide broad protection against discrimination of any kind on the basis of "any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status."[241] It would therefore violate the ICCPR, for example, for a State Party to enforce religious hate speech laws only when members of certain religions are affected or to adopt laws that ban hate speech involving some religious groups but not others.

In his report to the U.N. General Assembly in 2019, the U.N. Special Rapporteur called on states to broaden the application of ICCPR Article 20 (which covers national, racial, and religious groups) to all the grounds covered in Articles 2 and 26.[242] This call could expand hate speech laws to cover many more categories of protected persons, thereby broadening Article 20's reach. However, the guarantee of equal protection of the law without discrimination could also be used to invalidate hate speech laws that engage in discriminatory coverage of protected groups (e.g., laws that cover gender but not political opinion). This is a topic to monitor for developments on hate speech in the U.N. system.

4. Observations

In sum, in reviewing the interpretations of U.N. experts with respect to mandatory and discretionary bans on hate speech, several conclusions are evident. First, the U.N. machinery's interpretations take a narrow approach to the potential breadth of ICCPR Article 20 and CERD Article 4 by recommending strict definitions of key terms. For example, the Special Rapporteur has recommended that ICCPR Article 20 be interpreted to require the speaker have an intent to incite particular harm and that the harm

240. GC 34, *supra* note 128, at ¶ 26 (noting speech restrictions "must not violate the non-discrimination provisions of the Covenant").

241. ICCPR, *supra* note 115, arts. 2, 26. ICCPR Article 2 protects individuals from governmental discrimination in the provision of the treaty's rights and Article 26 requires equal protection of all laws. *Id*.

242. *U.N. Special Rapporteur 2019 Report*, *supra* note 140, at ¶ 9.

is likely and imminent.[243] The CERD Committee has invoked similar thresholds for all prosecutions under Article 4.[244]

Second, the application of ICCPR Article 19(3)'s legality and necessity tests to all hate speech bans, including mandatory ones, further constrains the space for hate speech bans. The legality test's prohibition on vagueness and overbreadth has driven findings of invalidity of a range of hate speech prohibitions throughout the world.[245] The application of Article 19(3)'s necessity test has resulted in U.N. experts pressuring governments to seek non-censorial methods to promote tolerance.[246] When non-censorial methods are insufficient, the U.N. machinery has applied high thresholds for governments to demonstrate they have selected the least intrusive means, including by requiring that speakers have an intent to cause harm and that such harm is imminent and likely.[247] Altogether, a review of the U.N.'s approach to hate speech over the last decade reveals a principled and disciplined framework that narrows the potential breadth of hate speech laws.

III. COMPARISONS AND IMPLICATIONS

Part III(A) compares U.S. and U.N. approaches to freedom of expression. It finds that the two standards are more similar than is commonly understood to be the case. Part III(B) examines the implications of this comparison with respect to the debate about whether social media companies should align their speech codes with U.S. or U.N. standards.

A. Comparing the First Amendment and U.N. Standards

A review of U.S. and U.N. standards on freedom of expression reveals numerous similarities with respect to their foundational underpinnings. Both start with a presumption in favor of speech with the burden on the speech regulator to demonstrate that any speech restriction is justified. In the United States, the Supreme Court's jurisprudence reflects a presumption in favor of speech with the burden on the government to prove the validity of restrictions.[248] With respect to U.N. standards, that presumption in favor of speech

243. *See supra* notes 210–12 and accompanying text.
244. *See supra* notes 218–22 and accompanying text.
245. *See supra* notes 227–31 and accompanying text.
246. *See supra* notes 234–35 and accompanying text.
247. *See supra* notes 236–38 and accompanying text.
248. *See supra* notes 66–76 and accompanying text.

is set out in ICCPR Article 19 and reinforced by the U.N. machinery's interpretations.[249]

In addition to the presumption in favor of speech, both bodies of law require that restrictions not be improperly vague or overly broad. In First Amendment jurisprudence, this requirement was developed by case law.[250] In U.N. standards, this requirement emerges from the "provided by law" or legality prong of ICCPR Article 19(3)'s tripartite test.[251] Interpretations of both bodies of law note that this requirement is in place to give fair notice to individuals of prohibited speech, to prevent selective prosecution by governmental officials, and to avoid chilling speech.[252] The prohibition on vague or over broad speech bans has been a powerful doctrine for finding a wide variety of speech bans to be invalid under both bodies of law.[253]

Both systems of law require that the imposition of restrictions must serve important public interest objectives. Though First Amendment jurisprudence does not provide an exclusive listing of potential public interest objectives, the Supreme Court does require that governmental objectives be either compelling or substantial[254] and has rejected objectives that were pretexts for other motives.[255] With regard to the ICCPR, this requirement appears in the "legitimacy" prong of Article 19(3)'s tripartite test.[256] The U.N. machinery has asserted that the listed public interest objectives in ICCPR Article 19(3) should be interpreted strictly.[257] U.N. experts will not tolerate invocations of public interest objectives as pretexts for other motives and have required states to demonstrate their cited objectives are truly warranted, including through evidence-based justifications.[258]

---

249. *See supra* note 149 and accompanying text.

250. *See supra* notes 42–44 and accompanying text.

251. *See supra* notes 151–52 and accompanying text.

252. *See supra* notes 44, 151–52 and accompanying text.

253. *See supra* notes 45–58, 154–71 and accompanying text.

254. *See supra* notes 68, 76 and accompanying text.

255. *See supra* note 71 and accompanying text.

256. *See supra* note 175, and accompanying text. To an American-trained lawyer, the word "legitimate" may connote any purpose that is not illegitimate, or even an almost rubber stamp review of alleged governmental purposes. *See* CHEMERINSKY, *supra* note 22, at 734–35 (describing the U.S. constitutional "rational basis" test that applies when strict or intermediate scrutiny is not applicable). However, a review of U.N. experts' application of the legitimacy test reveals that they require such objectives to be important and press governments to justify their stated objectives for burdening speech. *See supra* notes 177–86 and accompanying text.

257. *See supra* note 176 and accompanying text.

258. *See supra* notes 177–78, 181 and accompanying text.

Both U.S. and U.N. standards also require narrowly tailoring speech restrictions to avoid burdening speech more than is necessary, which has resulted in the invalidity of numerous speech restrictions under both systems.[259] The U.S. system provides different levels of narrow tailoring depending upon the type of restriction that is at stake. For content-based restrictions, courts apply strict scrutiny, which means the government must prove it has a compelling interest to restrict speech and it has selected the least restrictive means to advance that objective.[260] If a restriction is content-neutral, then courts will apply intermediate scrutiny, which is not as exacting as strict scrutiny but requires a close fit between the restriction and a substantial governmental objective.[261]

With respect to the ICCPR, the narrow tailoring requirement derives from Article 19(3)'s necessity prong, which requires, *inter alia*, that any speech restrictions be the "least intrusive means" of achieving a specified important public interest objective, which involves the examination of various contextual factors.[262] Under applicable U.N. standards, a government bears the burden of showing there are no non-censorial means to achieve the public interest objective, the government has selected the least intrusive of the available measures to achieve the objective, and the selected measure is effective.[263] U.N. standards do not distinguish between content-based and content-neutral restrictions and therefore apply the high threshold of the "least intrusive means" test to all limitations on speech.

Both standards have identified categories of unprotected speech. In the U.S. system, the Supreme Court has made normative assessments that certain types of speech merit less protection (e.g., advocacy of incitement to imminent violence and lawless action, true threats, and obscenity).[264] In the U.N. system, the treaty texts, particularly ICCPR Article 20 (which bans advocacy of national, racial or religious hatred that incites certain harms) and CERD Article 4 (which bans various forms of racist hate speech), set forth unprotected categories of speech.[265]

At first glance, it may appear that this difference in unprotected speech categories, particularly the U.N. system's mandatory

259. *See supra* notes 69–73, 80–81, 192–201 and accompanying text.
260. *See supra* notes 67–68 and accompanying text.
261. *See supra* notes 75–76 and accompanying text.
262. *See supra* note 187 and accompanying text.
263. *See supra* note 189 and accompanying text.
264. *See supra* notes 82–84 and accompanying text.
265. *See supra* notes 207, 219 and accompanying text.

hate speech bans, may be where substantial differences arise between the two bodies of law. But further analysis reveals that the potential breadth of these mandatory bans is significantly constrained in the U.N. human rights machinery's interpretations for two reasons. First, in the last decade in particular, U.N. experts have issued narrow interpretations of these mandatory hate speech bans. U.N. expert mechanisms have required any restrictions imposed under ICCPR Article 20 or CERD Article 4 to meet high thresholds, including that the speaker has an intent to incite harm, the harm is likely, and the harm is imminent.[266] These factors require taking context (and not just content) into account in determining whether speech may be restricted under U.N. standards. Similarly, even though hate speech is not categorically excluded from First Amendment protection based solely on content (in contrast to obscenity), hate speech in particular contexts is unprotected, including under intentional incitement and true threats standards. Second, the fact that ICCPR Article 19's tripartite test applies to all speech restrictions, including mandatory hate speech bans, further reduces a potential gap between the U.S. and U.N. standards. For example, U.N. mechanisms have often deemed hate speech restrictions invalid on vagueness grounds (as has occurred in the U.S. legal system).[267] In addition, the requirement of narrowly tailoring restrictions through the least intrusive means test has similarly diminished the scope of hate speech bans in the U.N. system, particularly with respect to U.N. expert interpretations requiring an examination of contextual factors to determine likely and imminent harm to satisfy the necessity test.[268]

In sum, comparing the U.S. and U.N. approaches to freedom of expression reveals that they share four core doctrines that make them principled bodies of law for strongly protecting freedom of expression. First, they both enforce a presumption in favor of speech with the speech regulator bearing the burden to demonstrate that restrictions are valid. Second, both bar speech restrictions that are either unduly vague or over broad—two powerful checks on speech regulation. Third, both require significant governmental reasons for banning speech and dismiss pretextual or unimportant reasons for burdening speech. Fourth, both require narrow tailoring of speech restrictions, with the U.N. system requiring a "least intrusive means" test for all speech bans whereas the United States uses a "least restrictive means" test only for content-based

266. *See supra* notes 209–21 and accompanying text.
267. *See supra* notes 99–102, 227–31 and accompanying text.
268. *See supra* notes 236–39 and accompanying text.

restrictions. Although both systems appear to differ significantly on hate speech, those differences are narrowed substantially in the U.N. experts' strict interpretations of high thresholds for mandatory hate speech bans as well as the application of ICCPR Article 19(3)'s rigorous tripartite test to all hate speech restrictions. Similarly, the differences are narrowed by First Amendment jurisprudence that permits hate speech to be punished when it satisfies certain criteria such as intentional incitement to violence and lawless action as well as true threats.

While this Article does not take the position that the two bodies of law completely converge, it does argue that both are based on foundational doctrines that impose a disciplined, principled, and speech-protective framework on speech regulators. This examination of the U.S. and U.N. approaches to speech illuminates that using "outlier" to describe the U.S. approach to speech is a misnomer when comparing these two bodies of law. Rather, the domestic approaches of many countries and certain interpretations from regional human rights systems are inconsistent with contemporary interpretations of U.N. speech standards.

B. Implications for Social Media Content Moderation

This comparison of U.N. and U.S. approaches reveals several implications for the use of international human rights standards for content moderation on social media platforms. First, the argument that social media companies should not move towards U.N. standards because they are not speech protective appears overblown and based on assessments of U.N. standards that do not include key interpretations from the last decade. Both standards are grounded in similar doctrines that discipline the speech regulator when limiting speech, as discussed in Part III(A).

Second, an in-depth examination of the U.N. machinery's guidance displays that—despite concerns that U.N. standards do not provide sufficient guidance to social media companies[269]—there is a very significant body of interpretations involving freedom of expression, which will provide social media companies with useful guidance in applying U.N. speech norms in their content moderation practices. Such interpretations are publicly available and they set forth substantial guidance on crucial concepts, including vagueness, overbreadth, the application of the least intrusive means test, legiti-

269. *See supra* note 15 and accompanying text.

mate public interest objectives, and the meaning of incitement.[270] Though no body of law can provide easy and pre-packaged answers to all potential online speech scenarios, the U.N. texts and interpretations of the treaty monitoring bodies and Special Rapporteur provide principled guidance that can be used to guide private sector content moderation.

Third, if society's normative objective is for social media companies to tie themselves to a principled "mast" in making decisions that involve freedom of expression—rather than tying speech decisions to the winds of profit, politics, and public pressure—then U.N. treaties as interpreted by the relevant U.N. human rights machinery provide a useful framework for decision-making. As noted by the Special Rapporteur, translating U.N. speech standards into the context of private content moderation would mean companies bear the burden of demonstrating that (1) their speech restrictions are not improperly vague or overbroad and (2) the selected burdens on speech represent the least intrusive means (3) to advance public interest objectives.[271] In a world in which the private sector exercises enormous power over speech, it is normatively desirable that companies demonstrate adherence to this three-part test in order to develop a more principled and disciplined approach to content moderation. The Special Rapporteur has also highlighted that companies must be transparent with the public when they depart from U.N. standards in curating speech.[272] Moreover, it is evident from examination of the ICCPR's tripartite test that private companies which adhere to these standards would not need to adopt the same speech codes and risk homogenization of speech rules across platforms. Rather, platform speech codes could embody a variety of approaches to expression so long as they meet, among other things, the legality, legitimacy, and necessity tests.

---

270. *See supra* notes 151–242 and accompanying text. To the extent such critiques are based not on a lack of guidance from the UN's human rights machinery, but rather on purported inconsistencies within such guidance, I have previously addressed why such arguments are misplaced. *See* Aswad, *supra* note 130, at 618–43 (observing that arguments based on inappropriately collapsing U.N. and regional interpretations as well as disregarding the possibility of evolution in the U.N. human rights machinery's interpretations are not well founded).

271. *See Special Rapporteur 2018 Report*, *supra* note 10, at ¶¶ 26–29, 46–48 (describing how companies should apply ICCPR Article 19's tripartite test in content moderation); *Special Rapporteur 2019 Report*, *supra* note 140, at ¶¶ 46–52 (explaining how to translate U.N. standards into the context of content moderation).

272. *Special Rapporteur 2019 Report*, *supra* note 140, at ¶¶ 48, 51 (calling on companies to demonstrate publicly their compliance with U.N. standards and observing that companies should be transparent when they depart from these norms).

Fourth, the straightforward character of ICCPR Article 19's tripartite test may make it more feasible for social media companies to implement than First Amendment jurisprudence, which has differing rules and levels of scrutiny for content-based and content-neutral regulations as well as a variety of additional nuances that add to its complexity.[273] That said, as I have cautioned previously, the speech protections that are reflected in the current state of U.N. expert interpretations could be diluted.[274] For example, the private sector could (mis)interpret U.N. standards to serve corporate profit-seeking ends and thus develop a competing "jurisprudence" that negatively impacts the ongoing trajectory of developments in U.N. interpretations.[275] In addition, the United States has been a key player in protecting and strengthening free expression norms at the United Nations and its departure from relevant U.N. fora could result in the weakening of the norms.[276] Although the potential for such outcomes should be acknowledged as part of any discussion about using U.N. norms as the benchmark for content curation, such issues are usually eclipsed by debates grounded in long-standing assumptions of the vast differences between U.S. and U.N. free speech approaches.

CONCLUSION

One of the most significant challenges to human freedom in the digital age involves the sheer power of private companies over speech and the fact that power is untethered to existing free speech principles. In 2018, a number of stakeholders advocated that global social media companies align their terms of service with the U.N.'s human rights principles on freedom of expression. As the American legal academy and practitioners have generally taken the view that the United States is an outlier when it comes to free speech, concerns have been raised that aligning corporate speech codes with U.N. standards would diminish speech rights in a problematic way.

This Article examined the widely unchallenged assumption about the scope of the U.N.'s freedom of expression protections, particularly in light of the evolution of the U.N. machinery's interpretations since 2011. A methodical examination of contemporary U.N. standards demonstrates that key components which make the First Amendment a principled and disciplined body of law (i.e., the

273. *See, e.g.*, *supra* note 81 (describing limited public forums in First Amendment jurisprudence).

274. Aswad, *supra* note 9, at 63–64.

275. *Id.* at 64.

276. *Id.* at 63.

bar on vague or overbroad speech prohibitions, the requirement that government narrowly tailor restrictions to avoid burdening speech beyond what is necessary, the need for government to demonstrate appropriate public interest goals in restricting speech, and placing the burden on the speech regulator—rather than the speaker—to prove any restrictions on speech meet these conditions) are remarkably similar to U.N. standards. Indeed, the prevailing assumption of American exceptionalism on free speech is not accurate when one analyzes the last decade of the U.N. machinery's freedom of expression interpretations (though the designation of U.S. exceptionalism remains accurate when compared to the domestic and regional legal systems that do not comport with U.N. standards).

These results reveal that criticism of aligning U.S.-based social media companies' content moderation with U.N. standards rather than First Amendment standards is overblown given both standards are grounded in speech-protective principles. This debate about which standards to use in content moderation should be recalibrated to include contemporary interpretations of U.N. standards. Indeed, the comparison of First Amendment and U.N. standards reveals that the U.N. standards may be an even more workable (and yet speech-protective) set of principles to apply to the complex process of private sector content moderation over the speech of billions. Those concerned about the protection of free speech in the digital age may better spend their focus on the fact that private companies are generally untethered to any speech-protective doctrines rather than criticizing U.N. standards based on outdated misconceptions about global norms, leaving private sector content moderation at the mercy of profits, politics, and public pressure.

***