2021

United Nations Free Speech Standards as the Global Benchmark for Online Platforms' Hate Speech Policies

Nadine Strossen

UNITED NATIONS FREE SPEECH STANDARDS AS THE GLOBAL BENCHMARK FOR ONLINE PLATFORMS' HATE SPEECH POLICIES

Nadine Strossen[1]

In the United States and around the world, one of the most controversial forms of speech or expression is "hate speech" — a phrase that is not a legal term of art but is widely used to denote speech that conveys hateful or discriminatory views on bases such as race, ethnicity, religion, gender, and sexual orientation. Debates about hate speech restrictions rage globally in light of free speech concerns. The most consequential debate today concerns how dominant online platforms (the "Platforms") should define the hate speech that warrants removal or restriction under their content moderation policies. Given the unprecedented volume of communications posted to the Platforms, the Platforms remove or restrict hundreds of thousands of communications as proscribed hate speech on a daily basis. Nor does a day go by without complaints from Platform users across the ideological spectrum, alleging that the Platforms write and enforce their hate speech policies in an arbitrary and discriminatory manner.

In 2017, the U.S. Supreme Court noted that Platforms have become the most important forums for the exchange of information and ideas, including by and about public officials, and also about public affairs. As the Court has declared, such speech is "more than a matter of self-expression; it is the essence of self-government." In short, the Platforms have supplanted traditional government forums for the exercise of free speech, and the Platforms also exercise censorial power on a scale that, in the past, only governments have wielded. But the Platforms are not "state actors." Thus, the Platforms' content moderation policies are not constrained by the First Amendment's free speech guarantee, or any other checks that the U.S. Constitution places on governmental speech restrictions. To the contrary, in designing and implementing content

1. For outstanding editing assistance, Professor Strossen thanks NYLS student Kirsten Kovats. For their assistance with the footnotes, Strossen thanks NYLS students Kelly Barrett, Michael Kazamias, Kirsten Kovats, and Shukra Sabnis. For their editorial assistance, Strossen thanks *Michigan State International Law Review* members Jordan Amidon, Mary Bradley, Gary Cooper, Victoria Hansen, Bradley Harrah, and Mitchell Purdy. Strossen is above all indebted to the scholarship, encouragement, and assistance of Professor Evelyn Aswad. For their helpful reviews of earlier drafts of this Article, Strossen thanks Natalie Alkiviadou, Evelyn Aswad, Paul Coleman, Molly Land, Jacob Mchangama, and Richard Ashby Wilson.

moderation policies, Platforms are exercising their own First Amendment rights, which the U.S. government would violate if it acted to control those policies.

Alternative tools are required to rein in the Platforms' vast censorial power and to promote both individual free speech and democratic discourse. One proposal that recently has gained traction was advanced in 2018 by law professor David Kaye, then the United Nations Special Rapporteur on free speech. Invoking the U.N. 2011 Guiding Principles on Business and Human Rights, Kaye urged the Platforms to align their content moderation policies with the international human rights standards delineated in U.N. treaties and in relatively recent speech-protective reports that have been issued by U.N. officials and bodies authorized to interpret those treaties.

This Article builds on the analysis of Kaye and other experts who have endorsed this approach, based on speech-protective U.N. standards. The Article analyzes the pertinent U.N. treaty provisions and interpretive materials, and it shows that their key speech-protective principles dovetail with those in recent U.S. First Amendment law. It also explains why this approach is the most feasible for restraining the Platforms' vast censorial power and for facilitating the Platforms' resistance to government suppression pressures, while also respecting the Platforms' own free expression rights.

This Article recognizes that the complex body of U.N. treaty provisions and interpretations, as well as U.S. First Amendment caselaw, contain not only speech-protective elements, but also elements that reflect a narrower view of protected speech. This Article quotes experts who maintain that the overall trend in U.N. standards is toward increased speech protection; it also urges free speech proponents to promote the speech-protective aspects of the U.N. regime both within the U.N. system itself, and also as the basis for the Platforms' content moderation policies.

> [F]reedom of expression advocates should actively seek to entrench the victories that have been achieved both in the text of [certain United Nations (U.N.) treaty provisions] and in the most recent interpretations by the U.N. human rights machinery. These advocates should pursue an active strategy of stealing victory (broad speech protections) from the jaws of potential defeat (other U.N. treaty provisions and interpretations that are less speech-protective) by promoting consistent application of positive interpretations recommended by the U.N. machinery.
>
> — *Evelyn Aswad, Herman G. Kaiser Chair in International Law, University of Oklahoma College of Law (2020)*[2]

> United Nations entities should . . . encourage [social media] companies . . . [to] align their content policies on hate speech with international human rights norms and standards . . . [and to] develop tools that promote individual autonomy, security and free expression, and involve de-amplification, de-monetization, education, counter-speech, reporting, and training as alternatives, when appropriate, to the banning of accounts and the removal of content.
>
> — *U.N. Strategy and Plan of Action on Hate Speech, Detailed Guidance on Implementation for U.N. Field Presences (September 2020)*[3]

> [Under] [t]he international human rights framework . . . [nation] States should generally deploy tools . . . other than criminalization and prohibition, such as education, counter-speech and the promotion of pluralism, to address all kinds of hate speech.
>
> — *U.N. Special Rapporteur on Freedom of Speech, David Kaye (2019)*[4]

> In many countries, overbroad [hate speech] rules . . . are abused by the powerful to limit non-traditional, dissenting, critical, or minority voices, or discussion about challenging social issues. Hate speech . . . laws

---

2. Evelyn Mary Aswad, *To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?*, 77 WASH. & LEE L. REV. 609, 655–56 (2020) [hereinafter Aswad 2020].

3. United Nations Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nation Field Presences, at 25–39 (2020) [hereinafter Detailed Guidance].

4. David Kaye (Special Rapporteur), Promotion and Protection of the Right to Freedom of Opinion and Expression, ¶ 28, U.N. Doc. A/74/486 (2019) [hereinafter 2019 Special Rapporteur on Free Speech].

ironically are often employed to suppress the very minorities they purportedly are designed to protect.

— *U.N. Human Rights Committee (2016)*

I. INTRODUCTION

A major free speech issue, which long has been debated in the United States (U.S.) and around the world, and which has been the focus of recently burgeoning debates about online speech, is how to regulate "hate speech"— speech that conveys hateful or discriminatory views on the basis of identity factors such as race, ethnicity, religion, and sexual orientation[5] —consistent with applicable freedom of speech norms. The worldwide importance of this issue is underscored by the 2019 launch of the United Nations (U.N.) Strategy and Plan of Action on Hate Speech, announced by U.N. Secretary General Antonio Guterres.[6] This Article focuses on what is undoubtedly the most consequential issue in terms of which hate speech is permitted worldwide, and which is not. Specifically, given the outsized role that the dominant transnational online platforms (the "Platforms") play in our global communications systems, this supremely consequential issue is what content moderation standards Platforms should implement and enforce for hate speech.

To illustrate the preeminent practical consequence of this issue, consider just one statistic: in Facebook's most recent "Community Standards Enforcement Report," for the fourth quarter of 2020, it reported that *each day* it was removing, or otherwise "taking action on," approximately 437,000 posts on Facebook and Instagram that were deemed to constitute prohibited hate speech.[7] This number bears

---

5. "Hate speech" is not a legal term of art under either U.S. or international human rights law, nor is there any societal consensus about the term's specific meaning. The laws of various countries, as well as U.N. treaties, define the concept differently (often without using the term "hate speech"). In the U.S., there is no legal definition at all because Supreme Court Justices of all ideological stripes, since the 1960s, have consistently rejected the contention that speech conveying hateful or discriminatory messages should, for that reason, be excluded from First Amendment protection. *See, e.g.*, Matal v. Tam, 137 S. Ct. 1744, 1763 (2017) (reaffirming that "the public expression of ideas may not be prohibited merely because the ideas are themselves offensive to some of their hearers.") (citing Street v. New York, 394 U.S. 576, 592 (1969)). Throughout this article, the term "hate speech" is used to convey the meaning stated in the main text, because this is the general sense in which the term is used in applicable law and in everyday speech.

6. *See generally* United Nations Strategy and Plan of Action on Hate Speech (2019).

7. *See* Guy Rosen, *Community Standards Enforcement Report, Fourth Quarter 2020*, FACEBOOK (Feb. 11, 2021), https://about.fb.com/news/2021/02/community-standards-enforcement-report-q4-

repetition and reflection: *in just one single day, just one Platform, enforcing just one of many content moderation standards, removed or otherwise suppressed 437,000 communications*.[8] Moreover, the fourth quarter report announced (with apparent pride) that this number was significantly higher than the prior quarter's corresponding number of 357,000 — a jump of more than 22% — indicating that we can expect a continuing upward trajectory.[9]

II. THE PLATFORMS' REGULATION OF HATE SPEECH SHOULD BE BASED ON THE SPEECH-PROTECTIVE ELEMENTS OF U.N. FREE SPEECH STANDARDS

What is the best response to the singularly impactful and challenging question that this Article addresses — namely, what hate speech standards should Platforms enforce? This Article endorses the approach that has been forcefully advocated by prominent experts[10] in both freedom of speech and international law:[11] Platforms should adhere to the speech-

---

2020/#:~:text=On%20Facebook%20in%20Q4%20we,from%204%20million%20in%20Q3.

8. *Id.*

9. *Id.*

10. These experts include University of California, Irvine Law School Professor David Kaye, who is the immediate past U.N. Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression (the "Special Rapporteur on Free Speech"), having served in that role from 2014 through 2020. *Mr. David Kaye, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, U.N. HUM. RTS. OFF. HIGH COMM'R, https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/DavidKaye.aspx (last visited March 8, 2021). In 1993, the U.N. Commission on Human Rights (the "Commission") mandated that the Special Rapporteur on Free Speech, "make recommendations . . . on ways . . . to better promote and protect the right to freedom of opinion and expression," among other things. *Special Procedures of the Human Rights Council, Special Procedures*, U.N. HUM. RTS. OFF. HIGH COMM'R, www.ohchr.org/en/HRBodies/SP/Pages/Welcomepage.aspx (last visited Feb. 19, 2021).

11. Important advocates of this approach also include individual scholars whose work this Article cites, and the civil society organizations Access Now, Amnesty International, Article 19, and the Electronic Frontier Foundation. *See, e.g.*, Amnesty Int'l, *'Let Us Breathe!' Censorship and Criminalization of Online Expression in Viet Nam*, AI Index ASA 41/3243/2020 5, 58 (2020) (calling upon major online companies to "adopt new content moderation and community standards policies that are explicitly and primarily grounded in international human rights standards"); ARTICLE 19, SIDE-STEPPING RIGHTS:

protective free speech standards under international human rights law.[12] specifically, Platforms should adopt pertinent U.N. treaties and their authoritative interpretations by the U.N. bodies and officials who are responsible for recommending interpretations of these treaty provisions and monitoring their implementation.[13] This Article refers to the foregoing legal norms with terms such as "U.N. law" and "U.N. standards," neither of which are legal terms of art, for the sake of clarity and to accurately

---

REGULATING SPEECH BY CONTRACT 4–5 (2018), https://www.article19.org/resources/side-stepping-rights-regulating-speech-by-contract ("[S]ocial media companies . . . should ensure that their Terms of Service are . . . in line with international standards on freedom of expression.").

Other individuals and organizations have advocated that Platforms should abide by U.N. free speech (and other U.N. human rights) law in the face of governmental measures or requests that are inconsistent with U.N. norms but have not expressly addressed the related question of whether Platforms should align their content moderation policies with U.N. free speech standards. *See infra* text accompanying notes 71-73; *GNI Principles*, GLOB. NETWORK INITIATIVE, https://globalnetworkinitiative.org/gni-principles/ (last visited Feb. 19, 2021).

12. *See* David Kaye (Special Rapporteur), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, Hum. Rts. Council on Its Thirty-Eighth Session, U.N. Doc. A/HRC/38/35 (2018) [hereinafter 2018 Special Rapporteur on Free Speech]. In 2011, then-Special Rapporteur on Free Speech, Frank La Rue, issued a report recognizing that, "any restriction [on] the right to freedom of expression [on] the Internet must . . . comply with international human rights law." Frank La Rue (Special Rapporteur), *Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 14, U.N. Doc. A/66/290 (Aug. 10, 2011) [hereinafter 2011 Special Rapporteur on Free Speech]. However, the 2018 Special Rapporteur on Free Speech Report is the first such report that focused specifically on the responsibility of online intermediaries, including the Platforms, to respect U.N. human rights standards in their content moderation policies.

13. Kaye's 2018 report urging Platform compliance with "international human rights law" defined that term as "including the relevant [U.N.] treaties and interpretations of the treaty bodies and special procedure mandate holders and other experts, including the Rabat Plan of Action." 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 58(b). "Treaty bodies" include the Human Rights Committee and the Committee for the Elimination of Racial Discrimination, as discussed *infra* at note 110. "Special procedure mandate holders" include the Special Rapporteur on Free Speech. For discussion of the Rabat Plan of Action, *see infra* at text accompanying note 147; Evelyn Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 64 (2018) [hereinafter Aswad 2018] ("Given that an international human rights court solely dedicated to adjudicating rights [under the International Covenant on Civil and Political Rights] does not exist, the U.N. machinery's recommended interpretations of U.N. [free speech] standards have come primarily from the Human Rights Committee, the Special Rapporteur, and (occasionally) certain high-profile, non-binding consensus resolutions adopted by U.N. member states.").

describe and highlight the noteworthy fact that these legal norms emanate from U.N. treaties, officials, and bodies.

The pertinent U.N. treaties, as well as the multifarious documents through which U.N. officials and bodies interpret and implement them, contain strong speech-protective elements that overlap with key speech-protective elements contained in U.S. law. However, the U.N.'s complex body of material also contains elements that not only permit certain speech restrictions, but also mandate some hate speech restrictions. Therefore, in exhorting Platforms to align their content moderation policies with U.N. free speech law, Special Rapporteur on Free Speech Kaye referred in particular to the most speech-protective elements of that law; other proponents of this approach have done likewise.

Kaye and others forcefully argue that the overall recent trend in U.N. free speech law is toward increased speech protection,[14] and that more recent speech-protective interpretations should be understood to supersede earlier less-protective ones, as happens within U.S. free speech law.[15] Other experts point out that various U.N. officials and bodies continue to issue some free speech/hate speech-related documents that are inconsistent with the U.N.'s speech-protective materials.[16] In any event, it is critical for free speech proponents to exert all potential influence on U.N. agencies and officials — as well as on Platforms — to adhere to the important recent speech-protective interpretations within the U.N. system; this is the point of Professor Evelyn Aswad's opening epigram to this Article, and other experts have made the same point.[17] Because the U.N. standards are enforced in part by "states with weaker traditions of free speech and press and judicial independence,"[18] it behooves supporters of democracy and of

14. *See infra* text accompanying notes 35, 37, 103, and 105.

15. *See* Aswad 2020, *supra* note 2, at 637.

16. *See* MICHAEL FARRIS & PAUL COLEMAN, HERITAGE FOUND., FIRST PRINCIPLES ON HUMAN RIGHTS: FREEDOM OF SPEECH 4 (2020) ("[A]t the heart of the U.N. system is a contradictory — even schizophrenic — approach to freedom of expression, with both a pro-free speech and pro-censorship approach in existence at the same time.").

17. *See id.* at 27 ("If the pro-free speech 'side' is to take precedence [within the U.N. system], positive U.S. action is needed. Free speech must be . . . championed at the international level.").

18. Sarah H. Cleveland, *Hate Speech at Home and Abroad*, *in* THE FREE SPEECH CENTURY 210, 224 (Lee C. Bollinger & Geoffrey R. Stone eds., 2018).

free speech to advocate especially strong U.N. free speech protections as a prophylactic measure.[19]

Recognizing the complexities and inconsistencies within the body of U.N. free speech materials and interpretations, as well as the challenges of formulating specific content moderation policies that are consistent with the broad speech-protective elements of these U.N. standards, this Article focuses on one specific aspect of the important analytical work that is required to implement Special Rapporteur on Free Speech Kaye's approach.[20] Specifically, this Article aims to lay out the strongest plausible case for Platforms to align their content moderation policies with the most speech-protective U.N. free speech standards. Heeding the old adage that "the devil [or the angel] is in the details," this Article seeks to highlight the angelic details within the U.N. materials — from a free speech perspective — and to urge free speech proponents to promote these both in the U.N.'s ongoing work and in the Platforms' content moderation policies.

First, this Article explains the distinction between the global U.N. free speech regime and other bodies of free speech law that are often referred to as "international," noting the reasons why the Platforms should adhere specifically to the speech-protective elements of the U.N. free speech regime. Next, this Article outlines the advantages of this approach from the perspectives of both the Platforms' users and the Platforms themselves. This Article then explains the substantial overlap between key speech-protective principles under recent U.S. and U.N. law, which condone suppressing hate speech only in appropriately narrow circumstances, and which stress non-censorial approaches for reducing hateful attitudes and

---

19. *See id.* (explaining that, during treaty negotiations regarding U.N. treaty provisions on free speech/hate speech, the U.S. advocated for the provisions to provide more speech-protective standards than were then reflected in the Supreme Court's First Amendment jurisprudence, for precisely these reasons); *infra* note 143.

20. *See, e.g.*, Susan Benesch, *But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies*, 38 YALE J. ON REG. ONLINE BULL. 86, 110 (2020). *See also* Justitia, Future of Free Speech Project: Realizing a Human Rights Approach in the Era of "Platformization" (on file with author) (describing a planned report that will provide detailed guidance about content moderation policies for hate speech and mis- and dis-information that comply with U.N. free speech standards).

action, including "counterspeech"[21] measures such as education and dialogue.

As many expert U.N. officials and bodies have noted, hate speech restrictions that exceed the narrow contours permitted under the speech-protective elements in U.N. law too often "are used [by governments] to suppress critical or opposing voices,"[22] and are also used, "to the detriment of" racial and other minority groups.[23] Not surprisingly, the same problems afflict the Platforms' hate speech policies,[24] since the Platforms are now free to design and implement these policies in any way they choose, such as to promote their corporate self-interests and to capitulate to pressures from governments, including authoritarian governments.[25]

III. U.N. LAW IS DISTINCT FROM OTHER BODIES OF LAW THAT ARE SOMETIMES REFERRED TO AS "INTERNATIONAL": THE DOMESTIC LAW OF OTHER COUNTRIES; AND THE LAW UNDER REGIONAL HUMAN RIGHTS TREATIES AND THEIR INTERPRETATION BY REGIONAL HUMAN RIGHTS COURTS

As noted above, this Article specifically includes the term "U.N." when referring to its pertinent law and standards,[26] to underscore the important distinctions between U.N. law and other law that is sometimes labeled

21. While "counterspeech" is not an official legal term of art in either U.N. or U.S. law, it is widely used to embrace any expression that aims to "counter" or curb the ideas reflected in any controversial speech as well as the potential adverse impact of such speech. Advocacy of counterspeech, rather than censorship, reflects the substantiated conclusion that the most effective response to speech with controversial content, including hate speech, is not *less* speech, but *more*. *See* NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 156–82, 217–21 (2018).

22. Frank La Rue (Special Rapporteur), *Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 51, U.N. Doc. A/67/357 (2012) [hereinafter 2012 Special Rapporteur on Free Speech].

23. Comm. on the Elimination of Racial Discrimination, Gen. Recommendation No. 35, ¶ 20, U.N. Doc. CERD/C/GC/35 (2013) [hereinafter Gen. Recommendation 35].

24. *See e.g.*, STROSSEN, *supra* note 21, at 91–94.

25. To cite an example that occurred shortly before this Article was completed, on February 10, 2021, in response to demands from the Indian government, Twitter permanently blocked over 500 accounts that criticized the government for its conduct during farmers' protests. Karan Deep Singh, *Twitter Blocks Accounts in India as Modi Pressures Social Media*, N.Y. Times (Feb. 10, 2021), https://www.nytimes.com/2021/02/10/technology/india-twitter.html.

26. For further explanation of the use of this terminology see *supra* note 2 and text following note 14, *supra*.

"international" despite its non-global scope. This section elaborates upon those distinctions.

In my extensive experience discussing these issues with many diverse audiences — including audiences comprising many lawyers and law students — it is apparent that U.N. free speech law is not as well-known as it should be, considering its embodiment in treaties to which almost every country in the world is a party, giving it worldwide force.[27] My experience indicates that U.N. international human rights law tends to be less well-known than other bodies of human rights law that are also "international" in scope, insofar as they concern the law of other nations and hence are sometimes referred to as "international human rights law."[28] In contrast with these other bodies of law, only the U.N. regime has international support at the global level, transcending regional as well as national boundaries. For example, the major U.N. treaty governing freedom of speech is the International Covenant on Civil and Political Rights ["ICCPR"], to which 178 countries are parties (out of the 193 countries that are U.N. members).[29] Given the Platforms' global nature, their content moderation policies should conform to the global human rights standards embodied in U.N. treaties and interpretive materials. Correspondingly, especially given the variation among regional human rights treaties and the regional human rights court rulings interpreting

---

27. After my presentation about this topic at the in-person Symposium that led to this present Michigan State *International Law Review* Symposium, Professor Erik Bleich asked audience members to raise their hands if they had any knowledge of U.N. human rights treaties, and few did. Colleagues who also speak about this theme have reported similar experiences.

28. Examples of more broadly known bodies of human rights law include the domestic law of countries other than the U.S. and the law under regional human rights treaties, such as the European Convention on Human Rights, as interpreted by regional human rights courts, such as the European Court of Human Rights.

29. *See Member States*, U.N., https://www.un.org/en/member-states/index.html (last visited Feb. 19, 2020); *Non-member States*, U.N., https://www.un.org/en/sections/member-states/non-member-states/index.html (last visited Feb. 19, 2021). The 15 U.N. member states that are not parties to ICCPR are: Bhutan; Brunei Darussalam; Kiribati; Malaysia; Micronesia (Federated States of); Myanmar; Oman; Saint Kitts and Nevis; Saudi Arabia; Singapore; Solomon Islands; South Sudan; Tonga; Tuvalu; and United Arab Emirates. Moreover, one non-member observer to the U.N. *is* a party to ICCPR: the State of Palestine. Office of the High Commissioner for Human Rights, *Status of Ratification Interactive Dashboard*, U.N. (Mar. 15, 2021), https://indicators.ohchr.org; *Member States*, U.N. https://www.un.org/en/member-states/index.html (last visited Mar. 15, 2021).

them, it is inappropriate for global companies to adhere to any one body of regional human rights law. As Professor Aswad wrote:

> What basis does Twitter have for favoring (or applying) Europe's regional approach to human rights in its global operations over other regions' human rights instruments? It is only by citing to universal standards embodied in international human rights law that Twitter can claim to ground its worldwide rules in a fair manner.[30]

Those unfamiliar with U.N. free speech law are typically surprised to learn that it now contains key elements that are highly speech-protective, consistent with important speech-protective precepts of U.S. First Amendment law in general, and that it also accords substantial protection to much hate speech in particular. To be sure, as with any complex body of law that has evolved over time and is subject to interpretation by various individuals and bodies, certain interpretations and applications of U.N. free speech standards are in tension, or even inconsistent, with others.[31] Likewise, despite U.S. free speech law's generally speech-protective thrust, it too contains significant speech-suppressive elements,[32] and even its speech-protective aspects have only recently predominated — considering that the First Amendment was added to the Constitution in 1791 — reflecting a constant tug of war between broader and narrower interpretations of the free speech guarantee.

Although U.N. free speech law similarly contains censorial elements,[33] some leading experts concur that the predominant trend in recent U.N. law

---

30. Aswad 2018, *supra* note 13, at 45.

31. *See* Farris & Coleman, *supra* note 16.

32. *See, e.g.*, Erwin Chemerinsky, *The Roberts Court and Freedom of Speech*, 63 FED. COMMC'N. L.J. 579, 585 (2011) (describing multiple cases in which the Supreme Court, under the leadership of Chief Justice John Roberts, has rejected free speech claims and created new categorical exclusions from First Amendment protection for certain speakers, contrary to conventional wisdom that the Roberts Court has been consistently speech-protective).

33. As discussed *infra*, text accompanying notes 128–54, two key U.N. treaty provisions mandate restrictions on certain hate speech. Although important U.N. officials and bodies have strictly interpreted these provisions as permitting hate speech restrictions only in limited circumstances, it nonetheless remains the case that some U.N. bodies do not clearly, consistently adhere to these strict interpretations.

has been toward greater speech protection, including for hate speech.[34] Just as older U.S. Supreme Court ("Supreme Court") decisions are widely considered to have been effectively overruled by subsequent, more speech-protective decisions — even though the older rulings never have been officially overruled[35] — some experts maintain that older U.N. interpretations of treaty provisions regarding free speech/hate speech should also be considered to have been superseded by more recent speech-protective interpretations.[36] In any case, the many proponents of strong free speech protection around the world, including many human rights and digital rights activists and scholars, should recognize and encourage the speech-protective elements of U.N. law — while critiquing inconsistent elements — just as they do regarding other pertinent bodies of free speech law, including First Amendment law. Free speech proponents should also engage more proactively and consistently with the U.N. human rights

34. *See, e.g.*, Aswad 2020, *supra* note 2, at 634, 643, 657; Cleveland, *supra* note 18, at 230–31; 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 28. Some other experts have advocated repealing or revising certain U.N. treaty provisions so that the treaty language more clearly restrains government power to restrict hate speech. Amal Clooney & Philippa Webb, *The Right to Insult in International Law*, 48 COLUM. HUM. RTS. L. REV. 1, 53 (2017) (advocating that U.N. treaty language should "prohibit speech only where it intentionally incites violence or a criminal offence that is likely to follow imminently . . . as a result of the speech"). *See also* Farris & Coleman, *supra* note 16. For a persuasive rejoinder to the Clooney and Webb piece, see Aswad 2020, *supra* note 2, in which Aswad endorses Clooney and Webb's goal of securely embedding strong speech-protective standards in U.N. law, but argues that the most effective strategy for doing so is to promote the recent speech-protective trend in U.N. officials' and agencies' interpretations of those treaty provisions.

35. For example, even though the Supreme Court has never overturned its decision in Beauharnais v. Illinois, 343 U.S. 250 (1952), in which the Court declined to strike down a state law regulating racist speech, that decision is generally considered to be inconsistent with subsequent Court rulings, including N.Y. Times v. Sullivan, 376 U.S. 254 (1964). *See* ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 695, 1402 (Rachel E. Barkow et al. eds., 6th ed. 2019). Likewise, although the Court has never overturned Chaplinsky v. New Hampshire, 315 U.S. 568 (1942), in which it upheld a "fighting words" conviction, that decision is generally considered to have been largely eviscerated by multiple subsequent Supreme Court rulings, including United States v. Stevens, 559 U.S. 460 (2010). *See* CHEMERINSKY, supra, at 1387. As yet another example, the Supreme Court's landmark decision in Brandenburg v. Ohio, 395 U.S. 444 (1969), is very much in tension with multiple earlier decisions that have not been overturned. *See* CHEMERINSKY, *supra*, at 1375–76.

36. *See* Aswad 2020, *supra* note 2, at 637–43.

machinery on free expression issues rather than solely or primarily reacting to U.N. actions and interpretations after the fact.[37]

Important aspects of U.N. law that strongly protect free speech in general, and hate speech in particular, distinguish it from other bodies of law that are sometimes referred to as "international," such as the domestic law of other countries and regional human rights law. Those other bodies of law are in significant respects less speech-protective than both First Amendment and U.N. free speech law.[38] This is one reason that Kaye and other experts have recommended that the Platforms align their content moderation policies specifically with the speech-protective aspects of U.N. free speech law, noting that it is not sufficient for the Platforms instead to model such policies on less speech-protective regional or national free speech law.[39] As noted above, these experts have also

---

37. *See id.* at 655–56; 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://www.lawfareblog.com/un-special-rapporteurs-latest-report-online-content-regulation-calls-human-rights-default ("[J]ust as the vaguely worded First Amendment has crystallized into more concrete rules, so too can international law. Compared with the First Amendment, international law on freedom of expression is young, having come into existence only since World War II, and active engagement with these norms . . . could facilitate its ongoing development.").

38. For an extensive discussion of the multiple respects in which U.N. law more strongly protects free speech than does regional human rights law, see Aswad 2018, *supra* note 13, at 45. For example, U.N. and U.S. law protects the following kinds of controversial speech that European law does not protect: blasphemy and the denial of certain historic atrocities. *Id.* A significant general respect in which European law is less speech-protective than U.N. and U.S. free speech law is the European law's deference to government. Specifically, the European Court of Human Rights invokes a "margin of appreciation," which constitutes a thumb on the scale in favor of government interests when balancing them against free speech rights. *Id.* In contrast, under U.N. and U.S. free speech law, the thumb is on the opposite side of the scale; speech restrictions are presumed impermissible, and the government has the burden of overcoming that presumption by demonstrating — among other things — that the restriction is necessary and the least restrictive alternative to promote the government's countervailing important public purpose. *See infra* at text accompanying notes 87 and 115.

39. *See* 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 26 (noting that the European Court of Human Rights has been less speech-protective, in notable respects, than U.N. free speech law, and concluding that "[r]egional human rights norms cannot . . . be invoked to justify departure from international human rights protections"); *accord*, Aswad 2020, *supra* note 2, at 636–37 (referring to Germany's strict Internet regulation law, the "NetzDG," as violating U.N. free speech standards, and rejecting Germany's reliance on the European Court of Human Rights' jurisprudence in defending that law on grounds

stressed, as an added rationale for this recommendation, that only global standards — not regional standards — can fairly be enforced at a global level.[40]

Indeed, some human rights advocates maintain that companies should also resist complying with local laws that violate U.N. free speech standards, and some Platforms have pledged to do so.[41] For example, Amnesty International endorsed and applied that approach to Facebook and Google, given their especially enormous size and power, in a report that it issued on November 30, 2020; in that same report, Amnesty International also noted the applicability of this approach to communications/technology companies generally:

> According to international human rights standards, Facebook and Google should respect freedom of expression in their content moderation decisions globally, regardless of the existence of local laws that muzzle freedom of expression. While companies sometimes point to the difficulties posed by conflicting obligations under local and international legal standards, they should be guided by the U.N. Guiding Principles on Business and Human Rights, which state: "The responsibility to respect human rights is a global standard of expected conduct for all business enterprises wherever they operate. It exists independently of States' abilities and/or willingness to fulfill their own human rights obligations . . . . And it exists over and above compliance with national laws and regulations protecting human rights."[42]

---

that, "invocation of a *regional* treaty cannot justify a country's violation of its *international* treaty obligations").

40. *See supra* at text accompanying note 31.

41. See text accompanying notes 71–73 *infra*, which discuss the Global Network Initiative, whose "[p]articipating companies will . . . work to protect the freedom of expression rights of users when confronted with government demands [and] laws . . . to suppress freedom of expression . . . in a manner inconsistent with internationally recognized laws and standards."

42. *See* Amnesty Int'l, *supra* note 11, at 6 (quoting Commentary to Principle 11 of the U.N. Guiding Principles on Business and Human Rights (UNGPs)).

IV. ADVANTAGES TO THE PLATFORMS AND TO THEIR USERS OF THE PLATFORMS' ADHERING TO SPEECH-PROTECTIVE U.N. FREE SPEECH STANDARDS

1. The Platforms now exercise vast censorial power, unconstrained by the limits on government

In a 2017 decision, the Supreme Court declared, "[w]hile in the past there may have been difficulty in identifying the most important places . . . for the exchange of views, today the answer is clear. It is cyberspace . . . and social media in particular."[43] The Platforms are the most essential forums for debate and discussion about public affairs and public officials among "We the People,"[44] and for us to engage with public officials and candidates for public office. These facts make it vital for our democracy to maintain the same "uninhibited, robust, and wide-open"[45] free speech in these new virtual venues that the Court has historically protected in traditional venues, proclaiming: "Speech concerning public affairs is more than self-expression; it is the essence of self-government."[46] Yet, the First Amendment free speech guarantee, far from *restraining* the Platforms' speech-restricting policies,[47] actually *protects* any such policies as an exercise of the Platforms' own free speech rights.[48]

In terms of the current factual and legal realities, we now face the worst of both worlds when it comes to the Platforms' censorial power over everyone else's speech: on the one hand, as the Supreme Court acknowledged, the Platforms wield censorial power of a magnitude that in

---

43. Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017).

44. U.S. CONST. pmbl.

45. N.Y. Times v. Sullivan, 376 U.S. 254, 270 (1964).

46. Garrison v. Louisiana, 379 U.S. 64, 75 (1964).

47. As private sector entities, the Platforms are not constrained by the First Amendment, which generally binds only government actors except in a few, limited circumstances. *See generally* CHEMERINSKY, *supra* note 35. Various scholars and litigators have argued that Platforms' content moderation policies should be viewed as constituting state action subject to U.S. constitutional constraints, on a range of theories. To date, of the many courts that have considered such arguments, none has accepted any of them. *See, e.g.*, Prager Univ. v. Google LLC, 951 F.3d 991 (9th Cir. 2020).

48. *See, e.g.*, Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241, 258 (1974) (upholding a newspaper's First Amendment right to decide what not to publish as well as what to publish).

the past only governments have exercised;[49] on the other hand, the First Amendment imposes no constraint on that non-governmental power. Not surprisingly, therefore, we have witnessed an outpouring of complaints from across the political and ideological spectrums that the Platforms are wielding their speech-suppressive power to silence or mute important messages of all stripes. Donald Trump and other conservatives complain that the Platforms' content moderation decisions reflect an anti-conservative bias.[50] Conversely, Democratic officials complain that, in response to those allegations, the Platforms bend over backward to avoid suppressing conservative messages that many Democrats and progressives maintain should be suppressed on various grounds, including that they constitute disinformation or hate speech.[51] Social justice activists, members of racial and other minority groups, and right-leaning individuals and groups alike are united in complaining that their messages are disproportionately silenced.[52] Not a day goes by without media reports of claims that the Platforms have inappropriately silenced — or failed to

---

49. *See* Richard Ashby Wilson & Molly Land, *Hate Speech on Social Media: Towards a Context-Specific Content Moderation Policy*, 52 CONN. L. REV. 1, 5 (2020) (footnotes omitted):

Governments are no longer the primary regulators of speech. Their regulatory capacity has been far outstripped by some of the largest companies in the world…, which together regulate the speech of 3.7 billion active social media users….In a reversal of the historic roles, private corporations have at times become the de facto regulators of government speech, as when Facebook banned the Commander-in-Chief of Myanmar's military from the platform and removed over 400 other news, entertainment, and lifestyle pages linked to the military.

Even more dramatically, in January 2021, Facebook and other Platforms banned the Commander-in-Chief of the US. Billy Perrigo, *Facebook and Twitter Finally Locked Donald Trump's Accounts. Will They Ban Him Permanently?*, TIME (Jan. 7, 2021) https://time.com/5927398/facebook-twitter-trump-suspension-capitol/.

50. Shannon Bond, *Trump Accuses Social Media of Anti-Conservative Bias After Twitter Marks His Tweets*, NPR (May 27, 2020), https://www.npr.org/2020/05/27/863422722/trump-accuses-social-media-of-anti-conservative-bias-after-twitter-marks-his-twe.

51. Rachel Kraus, *Once Again, There Is No 'Anti-Conservative' Bias on Social Media*, MASHABLE (July 28, 2020), https://mashable.com/article/anti-conservative-bias-facebook/.

52. Rachel Elizabeth Cargle, *When White People Are Uncomfortable, Black People Are Silenced*, HARPER'S BAZAAR (Jan. 9, 2019), https://www.harpersbazaar.com/culture/politics/a25747603/silencing-black-voices/. *See also* Bond, *supra* note 50.

silence — significant messages about COVID-19, national and local elections, or other topics of public concern.[53]

The Platforms' unprecedented censorial power is especially problematic because their suppression decisions are not accompanied by the procedural protections that cabin government censorship, along with First Amendment constraints. For example, while Facebook — commendably — has chosen to permit some appeals from its decisions that bar speech or speakers, commentators have observed that Facebook nonetheless appears relatively immune to arguments about posts that it had removed under its hate speech policies.[54] A recent Facebook report showed that its takedowns of hate speech are among the most appealed."[55] This is not surprising given the inherent subjectivity in determining what constitutes hate speech. Yet these posts are also "the least likely to be restored."[56] For the period of October 2017 to March 2019, Facebook reported that it had considered 1.1 million appeals from hate speech takedowns but reversed itself in only 152,000 such cases.[57] In other words, in just eighteen months, nearly 1 million users proactively[58] but unsuccessfully contended that their posts were unjustifiably suppressed as hate speech by just one company.

In the third quarter of 2020, Facebook reported that it had restored fewer than 250,000 posts out of the more than 22 million that it had reportedly removed as prohibited hate speech, slightly more than one percent.[59] Facebook also reported that, due to the coronavirus, it relied mostly on automated takedown methods and did not permit users to appeal

---

53. *See, e.g.*, RECLAIM THE NET, https://reclaimthenet.org/ (last visited Feb. 19, 2021); Gennie Gebhart, *Who Has Your Back? Censorship Edition 2019*, ELEC. FRONTIER FOUND. (June 12, 2019), https://www.eff.org/de/wp/who-has-your-back-2019.

54. *See, e.g.*, Komali, Comment to *The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech*, HACKER NEWS (Sept. 11, 2017), https://news.ycombinator.com/item?id=15220305.

55. Jonathan Shieber, *Facebook Releases Community Standards Enforcement Report*, TECHCRUNCH (May 23, 2019), https://techcrunch.com/2019/05/23/facebook-releases-community-standards-enforcement-report/.

56. *Id.*

57. Guy Rosen, *An Update on How We Are Doing at Enforcing Our Community Standards*, FACEBOOK (May 23, 2019), https://about.fb.com/news/2019/05/enforcing-our-community-standards-3.

58. One can fairly assume that additional users believed that their posts were also unjustifiably suppressed as hate speech but did not undertake the effort to appeal.

59. Shieber, *supra* note 55; Rosen, *supra* note 57.

in most situations.[60] This practice substantially increased the number of users maintaining that their posts were wrongly taken down on hate speech grounds without recourse.[61] Automated takedown inevitably leads to more unjustified suppression under such an inherently context-dependent concept as hate speech.[62]

As powerful and influential as the Platforms are in the U.S., with the nation's relatively developed and diverse communications and information ecosystem, these Platforms wield even more power and influence in less developed countries, whose inhabitants have fewer means through which they can communicate and access information. For example, in some countries, Facebook effectively *is* the Internet for the many people who are dependent on mobile phone applications and for whom Facebook is the only accessible app.[63]

2. Almost all countries are parties to the U.N. treaties governing free speech and hate speech

In light of the extensive — and in some countries nearly exclusive — control that the Platforms wield over communications and information, if individuals are going to exercise meaningful freedom to convey and receive information and ideas, that freedom must be protected on the

---

60. *Transparency, Hate Speech*, FACEBOOK (February 2021), https://transparency.facebook.com/community-standards-enforcement#hate-speech ("NOTE: Due to a temporary reduction in our review capacity as a result of COVID-19, we could not always offer our users the option to appeal. We still gave people the option to tell us they disagreed with our decision, which helped us review many of these instances and restore content when appropriate.").

61. *Id.*

62. *See* Svea Windwehr & Jillian C. York, *Facebook's Most Recent Transparency Report Demonstrates Pitfalls of Automated Content Moderation*, ELEC. FRONTIER FOUND. (Oct. 8, 2020), https://www.eff.org/deeplinks/2020/10/facebooks-most-recent-transparency-report-demonstrates-pitfalls-automated-content:

Automated systems are simply not capable of consistently identifying content correctly. Human communication and interactions are complex, and automated tools misunderstand the political, social or interpersonal context of speech all the time. That is why it is crucial that algorithmic content moderation is supervised by human moderators and that users can contest takedowns.

63. Nick Farrell, *Developing Countries Think Facebook Is the Internet*, FUDZILLA (Feb. 10, 2015), https://www.fudzilla.com/news/36984-developing-countries-think-facebook-is-the-internet.

Platforms. Notwithstanding the different legal systems and standards among the world's countries, almost all of them have become parties to the two U.N. treaties that directly govern free speech in general and hate speech in particular: the ICCPR and the International Convention for the Elimination of All Forms of Racial Discrimination ["ICERD"], respectively.[64] As observed by Susan Benesch, executive director of the Dangerous Speech Project, "[n]o [other] source of rules for speech regulation is as widely known or formally adopted."[65] From the perspective of individuals worldwide, it would be advantageous for the Platforms — as well as government parties — to honor these U.N. treaties because they embody universal human rights, including free speech rights, regardless of one's identity, beliefs, or location.

As the next section of this Article details, the Platforms' adoption of and compliance with the speech-protective elements of U.N. free speech law not only would entail several principled advantages, but also would have strategic advantages from the perspectives of users, democratic governments, and the Platforms themselves. Additionally, such a move would have strategic advantages from the perspective of U.S. users in particular, as a more feasible and effective option than the Platforms' potential adherence to U.S. free speech standards.

---

64. Out of the 193 nations that are U.N. members, 183 are parties to ICERD, and 178 are parties to ICCPR. *See Member States*, *supra* note 31; *International Convention on the Elimination of All Forms of Racial Discrimination*, U.N. TREATY COLLECTION, https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-2&chapter=4&clang=_en (last visited Feb. 10, 2021); *International Covenant on Civil and Political Rights*, U.N. TREATY COLLECTION, https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-4&chapter=4&clang=_en (last visited Feb. 19, 2021). The ten U.N. member states that are not parties to ICERD are: Brunei Darussalam; Democratic People's Republic of Korea; Kiribati; Malaysia; Micronesia (Federated States of); Myanmar; Samoa; South Sudan; Tuvalu; Vanuatu. Moreover, two non-member observers to the UN *are* parties to ICERD: the Holy See and the State of Palestine. Office of the High Commissioner for Human Rights, *Status of Ratification Interactive Dashboard*, U.N. (Mar. 15, 2021), https://indicators.ohchr.org; *Member States*, U.N. https://www.un.org/en/member-states/index.html (last visited Mar. 15, 2021). The 15 member states that are not parties to ICCPR are listed in note 30 *supra*.

65. Benesch, *supra* note 20, at 89.

3. Strategic advantages of Platforms' adherence to speech-protective elements of U.N. free speech law

Adherence to a single set of standards would relieve or at least reduce the Platforms' burdens of attempting to comply with multiple different legal standards in different jurisdictions. Adherence to the speech-protective aspects of U.N. free speech standards would make it easier for Platforms to resist pressures from individual governments to restrict speech in ways that violate these standards.[66] After all, it bears repeating, almost every single country in the world is a party to the ICCPR and the ICERD. Therefore, by moderating content in accordance with those U.N. treaties — and in accordance with their speech-protective interpretations[67] — rather than in accordance with any inconsistent standards under any one country's laws, Platforms would be honoring each country's own international legal obligations. This point was forcefully made in the above-quoted November 2020 Amnesty International report. It strongly critiqued Facebook and Google for capitulating to the Vietnamese government's pressure to suppress speech that criticized its policies and advocated democratic reforms, in violation of U.N. free speech standards:

> Technology companies including Facebook and Google must urgently overhaul their content moderation policies to ensure that they are firmly grounded in international human rights standards . . . The reformulation of these policies will provide the companies with firm ground to stand on as they seek to resist the repressive censorship demands of the Vietnamese authorities and other governments around the world.[68]

Some Platforms have professed support for honoring U.N. free speech norms concerning content moderation, rather than any inconsistent domestic or regional law, in the face of government demands — even

---

66. *See* Molly Land, *Toward an International Law of the Internet*, 54 HARV. INT'L L.J. 393, 447, 452 (2013):
[T]echnology companies may welcome the normative guidance Article 19 offers. Technology companies today are constantly engaged in questions about how to balance their own terms of service, local law, commercial demands, corporate culture, and at times even their own sense of morality . . . Article 19 could also provide important political "cover" in negotiating the demands of local law.

67. *See infra* text accompanying notes 95-164.

68. Amnesty Int'l, *supra* note 11, at 8.

though they continue to flout such norms in practice, as illustrated by the just-discussed actions of Facebook and Google at the behest of the Vietnamese government.[69] Platforms must be pressed to go beyond mere lip service to U.N. free speech norms; rather, Platforms should consistently honor the U.N.'s strongest speech-protective norms.

One initiative that has sought to secure Platforms' compliance with speech-protective U.N. standards is the Global Network Initiative (GNI). The GNI was launched in 2008 to promote "internationally recognized rights to freedom of expression and privacy" as "enshrined in instruments such as the [ICCPR] and the U.N. Guiding Principles on Business and Human Rights."[70] Facebook, Google, and other Platforms are members of the GNI, since they "agreed that" U.N. human rights standards constitute "the foundation for good law and policy."[71] Concerning freedom of expression, the GNI Principles state in part:

> Participating companies will respect and work to protect the freedom of expression of their users by seeking to avoid or minimize the impact of government restrictions on freedom of expression . . .
>
> Participating companies will respect and work to protect the freedom of expression rights of users when confronted with government demands, laws and regulations to suppress freedom of expression . . . in a manner inconsistent with internationally recognized laws and standards . . .
>
> When faced with a government restriction or demand that appears . . . inconsistent with . . . international human rights laws and standards on freedom of expression . . . participating companies will in appropriate cases . . . seek . . . modification from authorized officials of such requests; [s]eek the assistance, as needed, of . . . international human rights bodies

---

69. *Id.* at 6; Benesch, *supra* note 20, at 99 (referring to Facebook's self-proclaimed international human rights standards as part of its commitment to a "holistic and comprehensive" content moderation process, Benesch observes that, "[w]hile this public commitment is somewhat reassuring, accounting for human-rights standards as part of a holistic evaluation is very different than placing them at the core of your decision-making process").

70. *About GNI*, GLOB. NETWORK INITIATIVE, https://globalnetworkinitiative.org/about-gni/ (last visited Feb. 19, 2021).

71. *Id.*; *Our Members*, GLOB. NETWORK INITIATIVE, https://globalnetworkinitiative.org/#home-menu (last visited Feb. 28, 2021).

> or non-governmental organizations; and [c]hallenge the government in domestic courts.[72]

As another indication of some Platform support for the notion that content moderation policies should respect U.N. free speech norms, Facebook's quasi-independent Oversight Board, launched in 2020, has professed fidelity to such norms in exercising its authority to make final decisions regarding certain content moderation issues, stating: "[t]he Board will review whether content is consistent with . . . a commitment to upholding freedom of expression within the framework of international norms of human rights."[73] Indeed, the Board's first six decisions -- which it issued in January and February, 2021 — all cited the key U.N. treaty provision that protects free speech — Article 19 of the International Covenant on Civil and Political Rights (ICCPR) — and five of those six decisions reversed Facebook's removal of the speech at issue on the ground that the removal violated Article 19.[74]

Given liberal democratic countries' commitments to individual liberty, equality, and democracy, they should prefer content moderation standards that would facilitate the Platforms' resistance of censorial pressures from authoritarian governments. Indeed, as noted above, it behooves liberal democracies to promote, and to encourage Platforms to respect, the most speech-protective aspects of U.N. free speech principles precisely because these "principles would often be applied in states with weaker traditions of free speech and press and judicial independence."[75] Conversely, recent evidence shows that when liberal democracies have instead required or pressured Platforms to restrict speech that should be protected under U.N. standards, these speech-suppressive measures are later copied by authoritarian countries with severe adverse impacts on their citizens' free speech rights, equality, and democracy. The Danish-based human rights organization, Justitia, has issued two reports documenting the negative

---

72. *GNI Principles*, *supra* note 11; *Implementation Guidelines*, GLOB. NETWORK INITIATIVE, https://globalnetworkinitiative.org/implementation-guidelines/ (last visited Feb. 19, 2021).

73. *Announcing First Members of Oversight Board*, OVERSIGHT BD. (May 6, 2020), www.oversightboard.com/news/announcing-the-first-members-of-the-oversight-board/.

74. *See Board Decisions*, OVERSIGHT BD., https://oversightboard.com/decision/ (last visited Mar. 8, 2021).

75. Cleveland, *supra* note 18, at 224.

impact that Germany's "NetzDG"[76] law has had in many countries around the world since being implemented in 2018.[77] Presumably, German officials who supported this law did not to intend for authoritarian regimes to copy it, but they should have anticipated this predictable result. As Justitia's 2020 report concluded:

> While it would be misleading to blame Germany for the draconian laws adopted in authoritarian states, the fact that the spread of illiberal norms based on the NetzDG precedent has continued unabated should give Germany, liberal democracies and the European Commission food for thought when it comes to countering illegal and undesirable online content . . .
>
> In a world where both online and offline speech is under systematic global attack, democracies have a special obligation to err on the side of free speech, rather than succumbing to the ever-present temptation of fighting illiberal ideas with illiberal laws. Once democracies cede the high ground, authoritarians will rush in, creating a regulatory race to the bottom. This entails severe and negative consequences for free speech, independent media, vibrant civil society and political pluralism, without which authoritarianism cannot be defeated, nor democracy protected.[78]

Even from the perspective of Platform users in the U.S., with its strong constitutional protection of free speech against government regulation, it would be strategically advantageous for the Platforms to model their content moderation policies on U.N. free speech law. U.S. users who prize freedom of speech presumably would welcome the positive impact that this approach would have on free speech worldwide. In contrast, if the Platforms modeled their policies on U.S. free speech law, then other countries would likely resist the enforcement within their borders of a free

---

76. This is an abbreviation of the term "Netzwerkdurchsetzunggesetz," which is commonly translated as the "Network Enforcement Act." *NetzDG Notifizierung*, FRAGDENSTAAT, https://fragdenstaat.de/anfrage/netzdg-notifizierung/ (last visited Feb. 19, 2021).

77. JACOB MCHANGAMA & JOELLE FISS, JUSTITIA, THE DIGITAL BERLIN WALL: HOW GERMANY (ACCIDENTALLY) CREATED A PROTOTYPE FOR GLOBAL ONLINE CENSORSHIP 6–16 (2019) [hereinafter JUSTITIA 2019]; JACOB MCHANGAMA & NATALIE ALKIVIADOU, JUSTITIA, THE DIGITAL BERLIN WALL: HOW GERMANY (ACCIDENTALLY) CREATED A PROTOTYPE FOR GLOBAL ONLINE CENSORSHIP – ACT TWO 9–20 (2020) [hereinafter JUSTITIA 2020].

78. JUSTITIA 2020, *supra* note 77 at 3, 21.

speech regime that is generally perceived as "exceptionalist" in its strong speech protection. Specifically, many other countries would likely continue to require or pressure the Platforms to suppress much speech that the First Amendment would protect. For this same reason, the Platforms, with their global clientele, could not reasonably be expected to choose to base their content moderation policies on U.S. free speech law. Indeed, even within the U.S., many users and advertisers would be less likely to avail themselves of Platforms that permitted the full range of controversial speech that the First Amendment protects. Therefore, consistent with legitimate business concerns, the Platforms have imposed many speech restrictions that U.S. law bars government from imposing.

Notwithstanding widespread assumptions about the exceptionally speech-protective nature of U.S. free speech law, careful comparison of the U.N. approach to that of the U.S. demonstrates that the two share more key elements than has generally been recognized. It should also be recalled that the Platforms are under no obligation to respect users' free speech rights, and do not now do so. Therefore, if the Platforms abided by the U.N.'s speech-protective norms, this would markedly improve the situation of U.S. users, to whom the Platforms now have no free speech duties.[79]

V. IMPORTANT SPEECH-PROTECTIVE PRINCIPLES OF U.N. LAW DOVETAIL WITH MAJOR U.S. FREE SPEECH PRINCIPLES

> "[T]he [U.S.'s] protection of freedom of speech has fundamentally informed the shape of international human rights treaty protections, including in the context of hate speech. It continues to exert a constraining influence on treaty interpretation and jurisprudence."
>
> *— Sarah Cleveland, former member, U.N. Human Rights Committee, 2014–18 (2018)*[80]

Conventional wisdom paints First Amendment free speech principles concerning hate speech as extraordinarily speech-protective, and thus

---

79. *See* Aswad 2018, *supra* note 13, at 59–60 ("[A] rigorous and good faith interpretation of ICCPR Article 19's tripartite test [for any valid speech restriction] would bring company speech codes much closer to First Amendment standards than what is currently happening with the curation of speech on platforms.").

80. Cleveland, *supra* note 18, at 230–31.

starkly distinguishable from corresponding internationally accepted free speech principles. This conventional wisdom is, in fact, wrong for two reasons: first, it overstates how speech-protective U.S. law is; and second, it understates how speech-protective U.N. free speech law is. Immediately below is a summary of these two flaws in the conventional wisdom, which the following sections then discuss in detail.

While U.S. law does bar government from restricting hate speech solely because of its hateful message, U.S. law permits government to restrict such speech in particular contexts when the speech directly causes, or threatens to cause, certain serious imminent harms. In other words, the U.S. government may enforce any speech restriction that is necessary to avert those harms. For example, if a hateful message, in a particular context, intentionally causes or threatens to cause imminent discriminatory or violent conduct, then it can be punished. Likewise, U.N. free speech law bars any hate speech restriction unless it is necessary to avert certain serious harms — most importantly, discriminatory or violent conduct. Moreover, under both U.S. and U.N. free speech law, the speech restriction must be "the least restrictive alternative."[81] If the harm could be averted through an alternative measure that is less speech-restrictive, then that alternative measure must be employed. In sum, under both U.S. and U.N. law, speech may be restricted only if, in a particular context, it causes or threatens a specific harm that can be averted only through the restriction at issue. Finally, under both U.S. and U.N. free speech law, any speech restriction must be written in "narrowly tailored" language, with sufficient clarity and precision to guide both individuals who are subject to it and officials who enforce it, to avoid deterring or punishing lawful speech.

To further demonstrate the significant convergences between U.S. and U.N. law concerning hate speech, this Article will first summarize the key speech-protective principles of U.S. First Amendment law, and then explain how they overlap with key U.N. speech-protective principles.

1. Key speech-protective First Amendment principles

First Amendment law consists of a complex body of principles and standards that the Supreme Court has forged in hundreds of cases over

81. Sometimes slightly different words are used to convey this concept: "less" instead of "least"; "intrusive" instead of "restrictive"; and "means" instead of "alternatives."

more than a century. Nonetheless, the current speech-protective nature of this law, which strictly limits government power to regulate controversial speech, such as hate speech, largely follows from three cardinal general[82] principles. Two of these general principles impose substantive limits on government power, and one constrains how government formulates any speech regulation. Each principle has been endorsed by every Supreme Court Justice in modern U.S. history, and two of the three principles are closely paralleled in U.N. free speech law. These principles, on which this Article elaborates below, are often summarized with the following shorthand labels: (1) the viewpoint (or content) neutrality requirement; (2) the emergency test or judicial strict scrutiny; and (3) the narrow tailoring requirement.

A. The viewpoint (or content) neutrality requirement

In a landmark 1989 decision, the Supreme Court hailed the viewpoint neutrality principle as "the bedrock" sustaining the complex edifice of modern, speech-protective First Amendment law.[83] Sometimes referred to as "content neutrality," this principle bars government from restricting speech solely due to the disfavored nature of its viewpoint or content — *i.e.*, its message or idea. No matter how deeply loathed a viewpoint might be by no matter how significant a portion of the public, this does not justify censoring it. Accordingly, even horrifically hateful expression, which is deeply detested by the overwhelming majority of the community, may not be suppressed solely on that ground.[84]

B. The emergency test/judicial strict scrutiny

The second cardinal principle in U.S. free speech law, which is sometimes called the "emergency" test, permits government to suppress speech for reasons that go beyond its content or viewpoint, by permitting the government to consider the overall context in which it is spoken. If, in particular facts and circumstances, speech directly causes or threatens certain imminent serious harm and the harm cannot be averted in any other

---

82. The adjective "general" signals that, as with any complex body of law, even fundamental principles are subject to exception. *See infra* notes 84–87 and accompanying text.

83. Texas v. Johnson, 491 U.S. 397, 414 (1989).

84. *See* STROSSEN, *supra* note 21, at xxiv-xxv.

way, then government may restrict the speech. In short, government may restrict the speech if the restriction is necessary to prevent an emergency because no alternative measure, which is less speech-restrictive, will suffice.[85]

The specific criteria in the First Amendment emergency test can also be enforced through another, more general, test that courts regularly apply when reviewing restrictions on "fundamental" rights, including freedom of speech. Under this "strict scrutiny" test, a reviewing court closely examines the rights-restricting measure and the government's asserted justifications for it. The restrictive measure is presumed unconstitutional, and government can overcome that presumption only by demonstrating that the measure is necessary to promote a goal of compelling importance, and the least restrictive means for doing so.[86]

Despite their different rubrics, both the emergency test and strict scrutiny entail the same essential elements. Under both, the restricted speech must be tightly and directly connected to some specific serious harm, such that the restriction is necessary and the least restrictive means to avert that harm. These tests are appropriately demanding, channeling the government's censorship power toward the particular speech that is most likely to cause serious harm. In contrast, these tests foreclose the government's censorship power that is most likely to cause serious harm; when government has discretion to restrict speech that lacks a tight and direct connection to a specific serious danger, that discretion can — and inevitably will — be used in an arbitrary or even discriminatory manner and as a pretext for targeting disfavored viewpoints or disempowered, marginalized speakers.[87]

---

85. *Id.* at xx-xxi.

86. CHEMERINSKY, *supra* note 35, at 588–89. Sometimes the least restrictive means test is described as requiring "narrow tailoring." *Id.* at 589. In this context, that term underscores that the government measure must be carefully designed to address the specific emergency and must not grant the government substantively broader authority beyond what is necessary to do so. *Id.* This use of the phrase "narrow tailoring" is distinct from its use to prescribe how the speech regulation must be formulated. *See infra* note 93.

87. *See* STROSSEN, *supra* note 21, at 81–94.

C. Examples of hate speech that may be restricted, consistent with the emergency/strict scrutiny requirements

Appropriately demanding as the emergency and strict scrutiny tests are, government can nonetheless satisfy them in multiple situations. The Supreme Court has laid out context-specific criteria for several categories of speech that government may restrict consistent with the emergency and strict scrutiny tests.[88] One such category is a "true threat."[89] The adjective "true" distinguishes the legal concept of a punishable threat from a broader category of expression to which we tend to apply the term "threat" in ordinary speech. In everyday parlance, people often use the term "threat" to refer to any expression that they find frightening or discomforting. In contrast, to be a punishable true threat, the speaker must direct the expression to a single individual or small group of individuals, and intend to instill a "reasonable" fear on the part of any such individual that s/he will be subject to harm. The "reasonable" fear requirement embodies an objective standard, not a subjective one. Therefore, the fact that the expression might frighten a particular individual, who is unusually sensitive, does not warrant treating it as a punishable true threat. Other examples of categories of speech that satisfy the emergency/strict scrutiny tests include intentional incitement of imminent violence that is likely to happen imminently[90] and targeted harassment or bullying.[91]

Under U.S. free speech law, hate speech may be, and has been, punished when it satisfies the foregoing specific, contextual standards for speech that may be restricted consistent with the general emergency and strict scrutiny tests. Thus, contrary to common misconceptions, U.S. free speech law does not accord *absolute* protection to *all* hate speech. Rather, U.S. free speech law draws a sensible line between protected and punishable hate speech — a line that is paralleled by speech-protective U.N. free speech norms, as explained below.

---

88. *Id.* at 59–66.
89. *Id.* at 60–62.
90. *Id.* at 62–63.
91. *Id.* at 64–65.

D. The narrow tailoring requirement for formulating speech restrictions, to avoid undue vagueness

The last of the three cardinal First Amendment principles that strictly limit government power to regulate controversial speech, including hate speech, is the requirement that any speech regulation must be "narrowly tailored" in how it is formulated,[92] using sufficiently clear and precise language. This principle is also reflected in U.N. free speech law.

Recognizing that no language is perfectly precise, and hence that all laws inevitably entail some vagueness, the narrow tailoring requirement bars only laws that are "unduly vague." Laws will be deemed to satisfy the narrow tailoring requirement if they provide sufficient notice to the average person as to what conduct or expression is prohibited, and likewise provide sufficient constraints on enforcing authorities' discretion. When confronted with an unduly vague speech restriction, people will engage in unwarranted self-censorship, even of speech that is lawful and constitutionally protected, to avoid possibly violating the restriction. Correspondingly, laws that are unduly vague vest enforcing officials with excessive discretion that they will predictably exercise in accordance with their own subjective values, or those of powerful community factions. At best, officials will enforce such laws in an unpredictable, arbitrary and capricious manner; at worst, officials will enforce such laws in a discriminatory manner, selectively suppressing disfavored viewpoints and speakers, and thus violating the "bedrock" viewpoint neutrality principle.[93]

---

92. As noted *supra* note 87, the phrase "narrowly tailored" is also sometimes used as a synonym for the least restrictive alternative test: the substantive requirement that any speech restriction must be necessary to advance the government's purpose and less speech-restrictive than other available measures for doing so. *See* STROSSEN, *supra* note 21, at xx-xxi.

93. *See* STROSSEN, *supra* note 21, at 69–71; *see also supra* note 83 (regarding Supreme Court's reference to the "bedrock" nature of this principle).

2. Key speech-protective U.N. free speech principles

A. Overview of the three pertinent U.N. treaty provisions and their interrelationship

The U.N. free speech law provision that is the counterpart to the First Amendment's free speech clause is ICCPR[94] Article 19. Subsections (1) and (2) of Article 19 lay out the general scope of free speech protection, while subsection (3) sets forth the general requirements for permissible speech restrictions.[95] Article 19(3)'s general requirements for permissible speech restrictions have been determined to constrain hate speech regulations in particular.[96] Both ICCPR Article 20(2) and ICERD Article 4 mandate certain restrictions on some hate speech.[97] Both restrictions

---

94. For the reader's convenience, this is a reminder that ICCPR and ICERD are abbreviations for the International Covenant on Civil and Political Rights and the International Convention for the Elimination of All Forms of Racial Discrimination, respectively.

95. G.A. Res. 2200A (XXI), International Covenant on Civil and Political Rights, art. 19(1)–(2) [hereinafter ICCPR].

96. *Id.* art. 19(3).

97. *Id.* art. 20(2); G.A. Res. 2106 (XX), International Convention on the Elimination of All Forms of Racial Discrimination, art. IV (Dec. 21, 1965) [hereinafter ICERD]. The U.N. Convention on the Prevention and Punishment of the Crime of Genocide, adopted in 1948, also requires parties to "punish" a narrowly defined category of hate speech, namely, "direct and public incitement to commit genocide." G.A. Res. 260 A (III), Convention on the Prevention and Punishment of the Crime of Genocide, art. III(a) (Dec. 9, 1948). Notably, the Soviet Union had actively promoted adding a provision to the Genocide Convention that would have required punishment of a much broader, vaguer category of hate speech: "public propaganda tending by its systematic and hateful character to provoke genocide," or "to make [genocide] appear as a necessary, legitimate or excusable act." *See* Cleveland, *supra* note 18, at 216.

For a detailed discussion of incitement to genocide, see RICHARD ASHBY WILSON, INCITEMENT ON TRIAL: PROSECUTING INTERNATIONAL SPEECH CRIMES *passim* (Cambridge University Press 2017). As Wilson explains, the pertinent "speech acts" sufficient to constitute incitement to genocide are "generally far removed from the typical instances of threatening or denigrating speech found in national criminal courts." *Id.* Underscoring the narrowness of such "international speech crimes," he explains that:

> Speakers charged with [such] crimes have generally articulated the most extreme animus and conscious intent to harm, and gone beyond . . . insult . . . and slander to incite others to commit mass atrocities. Moreover, their utterances usually

were strongly opposed by the U.S. and other Western democracies, but supported by the former Soviet Union and Eastern European Communist bloc nations, along with the global South.[98]

The ICERD was adopted in 1965, and the ICCPR in 1966. Especially during their earlier years, there was serious debate about the interrelationship among their three key provisions as applied to hate speech: ICCPR 19(3) and 20(2), and ICERD 4.[99] In more recent years, though, the U.N. bodies and officials charged with recommending interpretations of these provisions, and with monitoring their implementation,[100] have repeatedly emphasized that *any* speech restriction — including any hate speech restriction — must comply with ICCPR 19(3)'s strict requirements; this is true specifically for any hate speech restrictions that ICCPR 20(2) and ICERD 4 mandate.[101] In short, these important interpretive reports have insisted that *the less speech-protective language* of ICCPR 20(2) and ICERD 4 must be read in conjunction with

---

> occur in a context of an armed conflict, genocide and a widespread or systematic attack on a civil population.

Id. at 2.

In addition to the extremely narrow contours of this punishable category of hate speech, it is further limited by the requirements of ICCPR Article 19(3), discussed *infra* text accompanying notes 106–27. *See* 2011 Special Rapporteur on Free Speech, *supra* note 12 ¶¶ 37, 81; *accord* 2012 Special Rapporteur on Free Speech, *supra* note 22, ¶ 41.

98. Jacob Mchangama, *The Sordid Origin of Hate-Speech Laws*, 170 HOOVER INST. POL'Y REV. 45, 53 (2011) ("Clearly, most contemporary proponents of hate-speech laws do not share the same ideologies and methods as the communist states of the day. Yet they seldom mention or reflect upon the fact that such laws were proposed and advocated for by antidemocratic states in which freedom of expression (as well as all other basic human rights) was routinely violated. Nor do they mention that these states, often totalitarian, had a clear interest in legitimizing and justifying their repression with the use of human rights language.").

99. *See* ICCPR, *supra* note 95, arts. 19(3), 20(2); ICERD, *supra* note 97. For the full texts of these provisions, see *infra* text accompanying note 106 (ICCPR 13); text accompanying note 140 (ICCPR 20(2)); and note 150 (ICERD 4).

100. *See* Aswad 2018, *supra* note 13, at 64 ("Given that an international human rights court solely dedicated to adjudicating ICCPR rights does not exist, the U.N. machinery's recommended interpretations of U.N. [free speech] standards have come primarily from the Human Rights Committee, the Special Rapporteur, and (occasionally) certain high-profile, non-binding consensus resolutions adopted by U.N. member states.").

101. The U.N. Human Rights Committee (HRC) reached this conclusion at least as far back as 1997, stating that "restrictions on expression which may fall within the scope of Article 20 must also be permissible under Article 19, paragraph 3." Ross v. Canada, U.N. Doc. CCPR/C/70/D/736/1997, Judgment U.N. Hum. Rts. Comm., at 14, (2000).

— and *is subordinate to — the more speech-protective language* of ICCPR 19(3).

Although various U.N. officials and bodies continue to issue reports that do not clearly, consistently adhere to ICCPR 19(3)'s strict requirements, notable experts construe the overall historical pattern throughout the U.N. regime as reflecting "an evolution in the U.N. machinery's interpretations toward greater speech protections,"[102] paralleling the same evolution of U.S. First Amendment law. In fact, at the time that the U.N. General Assembly adopted both the ICERD and the ICCPR, the Supreme Court was still in the process of building its current speech-protective jurisprudence, and many speech-suppressive precedents remained the law of the land.[103] Accordingly, when U.N. officials and bodies later began to more consistently construe the key U.N. treaty provisions in a more speech-protective way, they were following the same general trajectory as the Supreme Court.

In a 2019 report, the Special Rapporteur on Free Speech provided the following "summary of [U.N.] instruments on hate speech," stressing the speech-protective trend, specifically toward hate speech, that is reflected in recent and current U.N. free speech law:

> The international human rights framework has evolved in recent years to rationalize what appear, on the surface, to be competing norms. [R]estrictions on . . . freedom of expression must be exceptional, and the State bears the burden of demonstrating the consistency of such restrictions with international law; prohibitions under [ICCPR] Article 20 . . . and [ICERD] Article 4 . . . must be subject to the strict and narrow conditions established under [ICCPR] Article 19(3) . . ., and States should generally deploy tools . . . other than criminalization and prohibition, such as education, counter-speech and the promotion of pluralism, to address all kinds of hate speech.[104]

B. ICCPR Article 19

As the foregoing overview explains, ICCPR Article 19 is the key provision that governs all speech restrictions under U.N. law. It provides:

---

102. Aswad 2020, *supra* note 2, at 634.
103. *See infra* note 143.
104. 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 28.

> 1. Everyone shall have the right to hold opinions without interference.
>
> 2. Everyone shall have the right to freedom of expression; this right shall include 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.
>
> 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary: (a) 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.[105]

For any speech restriction to pass muster under ICCPR 19(3), the relevant government must demonstrate that it satisfies all three elements of that provision's three-part test. These three elements are usually labeled "legitimacy," "legality," and "necessity." Their meanings are summarized as follows:

> 1. Legitimacy. This term signifies that any speech restriction must be designed to promote a purpose that is considered legitimate or permissible: specifically, one of the important public purposes that ICCPR 19(3)(a) and (b) enumerate: "the rights or reputations of others," "national security," "public order (ordre public)[106]" and "public health or

---

105. ICCPR, *supra* note 95, art. 19.

106. *See* Letter from Brian Burdekin, Fed. Hum. Rts. Comm'r, Hum. Rts. & Equal Opportunity Comm'n, to Nick Bolkus, Minister for Admin. Servs., Parliament (May 20, 1991) (addressing initial submission on proposed ban on political advertising), http://www.humanrights.gov.au/right-freedom-information-opinion-and-expression-0#Submissions.

> [The term "ordre public"] is clearly wider than the concept of 'public order' in the sense usually understood in Anglo-Australian law (dealing with prevention of breaches of the peace, offensive behaviour etc). It extends to the sum of rules which ensure the functioning of society or the set of fundamental principles on which society is founded. It equates with the 'police power' in United States jurisprudence, permitting regulation in the interests of legitimate public purposes. This power must itself, however be exercised in a manner consistent with human rights.

Id. at 12.

morals."[107] Moreover, these purposes may not be invoked as mere pretexts; governments must show that any speech restriction is designed to, and does, promote one or more of them.[108] Although the specified legitimate purposes seem quite expansive and malleable, they have been interpreted relatively narrowly by the U.N. Human Rights Committee (HRC)[109] and the Special Rapporteur on Free Speech. For example, the HRC has clarified that the "rights . . . of others" refers specifically to the rights protected under the ICCPR and other elements of U.N. human rights law.[110] Likewise, while "public morals" is a potentially elastic concept, the HRC and Special Rapporteur on Free Speech have stressed that it must be determined according to the "universality of human rights and the principle of non-discrimination," and must not "be based on principles . . . deriving exclusively from a single tradition."[111]

2. Legality. This term refers to the requirement that any restriction must be "provided by law," which has been construed to mean that the law must be "adopted by regular legal processes," and written "with sufficient precision."[112] These requirements serve to "limit government discretion" in enforcing such a restriction, while also providing adequate guidance to individuals subject to it.[113]

3. Necessity. This term refers to the requirement that any restriction must be "necessary," and the least restrictive alternative, to promote one or more of the enumerated legitimate purposes. U.N. officials and bodies with authority to interpret ICCPR 19 sometimes use the term

---

107. *See* Gen. Comment No. 34 on Article 19: Freedoms of Opinion and Expression, Hum. Rts. Comm. on Its One Hundred-Second Session, U.N. Doc. CCPR/C/GC/34, ¶ 22 (2011) [hereinafter Gen. Comment 34] (stating that "[r]estrictions are not allowed on grounds not specified" in ICCPR 19(3), "even if such grounds would justify restrictions to other rights protected" under the ICCPR).

108. *Id.* ¶ 30.

109. The HRC is the body of independent experts elected by the ICCPR's state parties — the nations that have explicitly consented to be bound by the ICCPR — with authority to monitor the state parties' compliance with the ICCPR and to recommend interpretations of it. *Human Rights Committee*, U.N. HUM. RTS. OFF. HIGH COMM'R, http://www.ohchr.org/en/HRBodies/ccpr/pages/ccprindex.aspx (last visited Feb. 19, 2021).

110. *See* Gen. Comment 34, *supra* note 107.

111. *Id.* ¶ 28; *accord* 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 7.

112. 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 7.

113. *Id.*

"proportionality" interchangeably, or in tandem with, the term "necessity."[114]

The foregoing three requisite elements for a permissible speech restriction under U.N. law expressly correspond to two of the three key requirements for a permissible speech restriction under U.S. law, as explained above.[115] The U.N. legitimacy requirement parallels the U.S. requirement that the law must have an important purpose (one element of the strict scrutiny/emergency tests). The U.N. legality requirement parallels the U.S. requirement that the law must be written in a narrowly tailored fashion, to avoid undue vagueness. Finally, the U.N. necessity requirement parallels the U.S. requirement that the law must be necessary to promote its important purpose, and the least speech-restrictive means for doing so (the second element of the strict scrutiny/emergency tests).

C. General Comment 34

The Human Rights Committee (HRC) is the body of independent experts, elected by the ICCPR's state parties, with authority to monitor compliance with the ICCPR and to recommend interpretations of it.[116] The HRC most comprehensively laid out its strict construction of ICCPR

---

114. *See, e.g.*, Gen. Comment 34, *supra* note 107, ¶ 34; Gen. Comment No. 27, Hum. Rts. Comm. on Its Sixty-Seventh Session, U.N. Doc. CCPR/C/21/Rev.1/Add.9, ¶ 14 (1999) [hereinafter Gen. Comment 27].

115. *See supra* at text accompanying notes 84-94. While ICCPR 19(3)'s three prerequisites for speech restrictions do not *explicitly* affirm the essential viewpoint/content neutrality principle in U.S. law, it can plausibly be argued that the legitimacy requirement does so *implicitly*. Consistent with that requirement, any speech-restricting measure must actually be designed to promote an important public purpose that is among the limited set of such purposes enumerated in 19(3). A viewpoint/content-based measure, by definition, restricts speech solely to suppress its disfavored viewpoint or content; that purpose is not even legitimate, let alone one of the enumerated important public purposes in 19(3). Therefore, a viewpoint-based restriction would necessarily violate not only the viewpoint/content neutrality principle in U.S. law, but also 19(3)'s legitimacy requirement in U.N. law.

Even beyond the fact that the viewpoint/content neutrality principle is implicitly reflected in 19(3), U.N. reports have also expressly construed the pertinent U.N. materials and interpretations as reflecting this principle. For example, in his 2012 report on hate speech, Special Rapporteur on Free Speech Frank La Rue said, "international law prohibits some forms of speech for their consequences, and not for their content as such." 2012 Special Rapporteur on Free Speech, *supra* note 22, at ¶ 46.

116. *Human Rights Committee*, *supra* note 109.

19(3)'s prerequisites for permissible speech restrictions in its General Comment 34, issued in 2011, following years of study, debate, and discussion among the HRC members, civil society, and nations around the world.[117] General Comment 34 also drew upon the analyses in multiple prior HRC reports, stressing that all three of Article 19(3)'s requirements are "strict,"[118] and that governments have a heavy burden of demonstrating that each one is satisfied in order to justify any speech restriction.[119]

General Comment 34 makes clear the many parallels between the speech-protective elements of more recent U.N. and U.S. free speech law. First, both bodies of law place the burden on the government to demonstrate that any speech-restricting measure complies with all prerequisites for permissibility,[120] and they both entitle the government to no deference regarding these determinations.[121] For example, it does not suffice for the government to assert a proper purpose for the measure; rather, the government must show that the measure actually is designed to promote such a purpose and indeed does so.[122] Moreover, in language highly resonant with the U.S. emergency and strict scrutiny tests, General Comment 34 states that, even when the government asserts a proper purpose for restricting speech — *i.e.*, a concern that the speech poses a particular threat to one of the important public interests enumerated in ICCPR 19(3)'s legitimacy requirement — the government still may not restrict the speech unless it can "demonstrate in specific and individualized fashion the precise nature of the threat, and the necessity and proportionality of the specific action taken, in particular by establishing a direct and immediate connection between the expression and the threat."[123]

---

117. Gen. Comment 34, *supra* note 107.

118. *Id.* ¶¶ 23, 30, 48.

119. *Id.* ¶ 27.

120. *Id.* ("It is for the State party to demonstrate the legal basis for any restrictions imposed on the freedom of expression.").

121. *Id.* ¶ 36 (reaffirming that government is not entitled to a "margin of appreciation" in assessing whether a speech restriction complies with ICCPR 19(3), unlike the less speech-protective enforcement scheme in European law).

122. *Id.* ¶ 22 ("Restrictions must be applied only for those purposes for which they were prescribed and must be directly related to the specific need on which they are predicated."). *See also id.* ¶ 36 ("[A] State party . . . must demonstrate in specific fashion the precise nature of the threat to any of the enumerated grounds . . . that has caused it to restrict freedom of expression.").

123. *Id.* ¶ 35.

In yet another key parallel with the U.S. emergency and strict scrutiny tests, General Comment 34 expressly incorporates the least restrictive alternative requirement in a rigorous, two-step fashion. First, it stipulates that governments may not restrict speech at all "if the protection" from the feared harm "could be achieved in other ways that do not restrict freedom of expression."[124] Second, even if the government shows that it is necessary to impose some restriction on speech, it must additionally show that the particular restriction at issue is "the least intrusive" option "amongst those which might achieve" the government's goal.[125]

Just as General Comment 34 closely corresponds to U.S. law's emergency and strict scrutiny tests, it also closely corresponds to U.S. law's requirement that a speech restriction must be formulated with "narrowly tailored" language to avoid undue vagueness. Specifically, General Comment 34 construed ICCPR 19(3)'s legality requirement (that any restriction must be "provided by law") to embody these same criteria, explaining that, "a norm, to be characterized as a 'law,' must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly . . . A law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution."[126]

D. The hate speech restrictions mandated in ICCPR 20(2) and ICERD 4 must comply with the strict limits that ICCPR 19(3) specifies for any speech restrictions

As indicated above, notwithstanding ICCPR 19(3)'s strong general protection of free speech and its strict general limits upon any speech restrictions, two significant provisions in U.N. treaties specifically focus on certain hate speech, expressly requiring governments to outlaw it: ICCPR 20(2) and ICERD 4.[127] While these two provisions certainly could be viewed as being in tension with ICCPR 19(3), the most recent

---

124. *Id.* ¶ 33.

125. *Id.* ¶ 34. *See also* 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 7 ("States 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. States may not merely assert necessity, but must demonstrate it, [both] in the adoption of restrictive legislation and the restriction of specific expression.").

126. Gen. Comment 34, *supra* note 107, ¶ 25.

127. *See* ICCPR, *supra* note 95, art. 20(2); ICERD, *supra* note 97.

interpretations by U.N. officials and bodies that expressly address this issue repeatedly have construed them as consistent with — specifically, as subject to — ICCPR 19(3).

Regarding ICCPR 20(2), General Comment 34 declared, "a limitation that is justified on the basis of [ICCPR] 20 must also comply with [ICCPR] 19, paragraph 3."[128] General Comment 34 subsequently reiterated and underscored this point: "In every case in which the State restricts freedom of expression it is necessary to justify the prohibitions . . . in strict conformity with [ICCPR] 19."[129] This conclusion has been reaffirmed in numerous U.N. documents and by various U.N. officials and agencies, including the Special Rapporteur on Free Speech and the HRC.[130]

The foregoing conclusions of General Comment 34 about ICCPR 20(2) were also endorsed concerning ICERD 4 by the Committee on the Elimination of Racial Discrimination (CERD), the expert body appointed to interpret and monitor compliance with ICERD.[131] CERD's 2013 General Recommendation 35, entitled "Combating Racist Hate Speech," cited the ICCPR 19(3) and General Comment 34, declaring that "freedom of expression . . . may . . . be subject to certain restrictions, but only if they are provided by law and are necessary for protection of the rights or

---

128. *See* Gen. Comment 34, *supra* note 107, ¶ 50 (citing Ross v. Canada, U.N. Doc. CCPR/C/70/D/736/1997, Judgment U.N. Hum. Rts. Comm. (2000)).

129. Gen. Comment 34, *supra* note 107, ¶ 52. *See also* 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 8 ("Restrictions pursuant to [ICCPR] 20(2) . . . must still satisfy the cumulative conditions of legality, necessity, and legitimacy."). General Comment 34 also explicitly rebuts a particular argument that the more specific provisions in ICCPR 20(2) and ICERD 4 should supersede the more general language in ICCPR 19(3). *Id.* ¶ 51. Specifically, the *lex specialis* interpretive principle stipulates that a law governing a specific subject matter presumptively overrides a law that only governs general matters. *Id.* As the pertinent text explains, "[F]or the acts addressed in [ICCPR] 20, the Covenant indicates the specific responses required from the State: their prohibition by law. It is only to this extent that [ICCPR] 20 may be considered as *lex specialis* with regard to [ICCPR] 19." *Id.*

130. *See, e.g.*, 2012 Special Rapporteur on Free Speech, *supra* note 22, ¶ 41; 2011 Special Rapporteur on Free Speech, *supra* note 12 ¶¶ 37, 81; *see also infra* note 137 (describing the HRC's 2016 ruling in Rabbae v. Netherlands).

131. *Committee on the Elimination of Racial Discrimination*, *Human Rights Bodies*, U.N. HUM. RTS. OFF. HIGH COMM'R, http://www.ohchr.org/en/hrbodies/cerd/pages/cerdindex.aspx (last visited Jan. 13, 2021). CERD is HRC's counterpart, occupying the same role vis-à-vis ICERD that HRC occupies vis-à-vis ICCPR.

reputations of others and for the protection of national security or of public order, or of public health or morals."[132]

As noted above, prominent experts have concluded that the U.N.'s free speech standards have evolved toward more speech protection. Those experts flag General Comment 34 and General Recommendation 35 as critical elements in this evolution, which are entitled to significant weight given the lengthy, deliberative processes that gave rise to them. Accordingly, such experts maintain that these general interpretive documents should be read as implicitly superseding prior inconsistent, less speech-protective U.N. documents that also construe the pertinent treaty provisions. For example, Professor Aswad views General Recommendation 35's narrow concept of punishable hate speech — one that integrates ICCPR 19(3)'s strict limits on any speech restriction — in this light:

> This General Recommendation should be viewed as an evolution in the [CERD] Committee's thinking on bans on racist hate speech rather than an inconsistency in its jurisprudence. Before issuing formal recommended interpretations of treaties, U.N. treaty committee members engage in a consultation process that canvases the views of State Parties and civil society organizations. After such consultations, committees hold intensive deliberations that can take years before issuing the committee's ultimate views. Such a period of intensive deliberation and reflection can easily mark a turning point and evolution in the U.N. CERD Committee's thinking on Article 4, just as seminal cases decided over a prolonged period can spark transitions in a domestic court's jurisprudence.[133]

As this Article's next section details, recent U.N. reports that specifically focus on hate speech, issued by multiple U.N. expert bodies and individuals, have carefully adhered to the speech-protective interpretations of the three pertinent treaty provisions in General Comment 34 (as echoed in CERD General Recommendation 35). These reports have stressed that government has only narrow authority to regulate hate speech, while urging greater reliance on non-censorial measures

---

132. Gen. Recommendation 35, *supra* note 23, ¶ 26. *See also id.* ¶ 12 (stating that "the application of criminal sanctions" to hate speech under ICERD 4 "should be governed by principles of legality, proportionality, and necessity," the three-part test delineated in ICCPR 19(3)).

133. Aswad 2020, *supra* note 2, at 639.

addressing the root causes of hateful and discriminatory attitudes, as well as measures that outlaw discriminatory or violent conduct.

For example, the U.N.'s most recent, most detailed report on point, its 2020 Detailed Guidance on Implementation of the U.N. Strategy and Plan of Action on Hate Speech (the "Detailed Guidance"), endorsed this speech-protective approach, including in the specific context of the Platforms' content moderation policies.[134] Moreover, the Special Rapporteur on Free Speech issued two reports, in 2018 and 2019, concluding that Platforms should align their content moderation policies with General Comment 34 and other speech-protective elements of U.N. free speech law. Both of these Special Rapporteur reports stress that hate speech should be restricted only in narrow, exceptional circumstances, while most hate speech, along with hateful attitudes and actions, should be addressed through non-censorial measures.[135] Additional important recent materials, issued by multiple U.N. bodies, have also adhered to this narrow concept of punishable hate speech.[136]

To be sure, not all U.N. reports that address free speech/hate speech issues have clearly and consistently adhered to the speech-protective interpretations of the treaty provisions that General Comment 34 spells out

---

134. *See infra* text accompanying notes 162-64.

135. *See infra* text accompanying notes 156–61.

136. *See* Cleveland, *supra* note 18, at 226 (noting that the HRC has issued only one decision addressing the merits of an ICCPR 20(2) claim, in which it held that the Netherlands' strict concept of punishable incitement, which was very similar to the strict U.S. standard under Brandenburg v. Ohio, 395 U.S. 444, 447–48 (1969), complied with ICCPR 20(2); therefore, Cleveland concluded that "this brings [ICCPR] 20 closer to *Brandenburg*"). *See also* Babbae v. Netherlands, U.N. Doc. CCPR/C/117/D/2124/2011, Judgment U.N. Hum. Rts. Comm. (2016); Aswad 2020, *supra* note 2, at 625 n.64 (describing Human Rights Council Resolution 16/18, adopted by consensus in 2011 to address intolerance based on religion or belief, and noting that it "call[s] upon states to ban offensive and hateful speech in only one instance – incitement to imminent violence"). Another important example, noted in Special Rapporteur on Free Speech David Kaye's 2018 report, is the Rabat Plan of Action. *See* 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 68.2. Adopted through an extensive consultative process involving forty-five experts from different cultural backgrounds and legal traditions and convened by the U.N. High Commissioner for Human Rights in 2011–2012, the Rabat Plan of Action detailed standards for permissible prohibitions on incitement to national, racial, or religious hatred consistent with ICCPR 19 and 20 and ICERD 4. *Id.*

and that General Recommendation 35 echoes.[137] Therefore, proponents of basing the Platforms' content moderation standards on speech-protective U.N. free speech norms must continue to promote those norms within the U.N. system, applauding U.N. reports that do so,[138] while critiquing those

---

137. The author of this article thanks Natalie Alkiviadou, Paul Coleman, and Jacob Mchangama for calling to her attention some recent HRC and CERD country reports that, in their view, do not evaluate countries' hate speech restrictions sufficiently strictly in accordance with those very bodies' general interpretations of the pertinent treaty provisions in General Comment 34 and General Recommendation 35, respectively. *See, e.g.*, Hum. Rts. Comm., Concluding Observations on the Fourth Periodic Rep. of Czechia, ¶¶ 16–17, U.N. Doc. CCPR/C/CZE/CO/4 (2019); Hum. Rts. Comm., Concluding Observations on the Sixth Periodic Rep. of Belgium, ¶¶ 19–20, U.N. Doc. CCPRC/BEL/CO/6 (2019); Comm. on the Elimination of Racial Discrimination, Concluding Observations on the Combined Twenty-First to Twenty-Third Periodic Rep. of the U.K. of Great Britain and Northern Ireland, ¶¶ 15–17, U.N. Doc. CERD/C/GBR/CO21-23 (2016); Comm. on the Elimination of Racial Discrimination, Concluding Observations on the Combined Nineteenth to Twenty-Second Periodic Reports of Germany, U.N. Doc. CERD/C/DEU/CO/19-22 (2015); Comm. on the Elimination of Racial Discrimination, Concluding Observations on the Combined Seventh to Ninth Periodic Reports of the United States of America, U.N. Doc. CERD/C/USA/CO7-9 (2014); Hum. Rts. Comm., Concluding Observations of the Sixth Periodic Rep. of Germany, ¶ 18, U.N. Doc. CCPR/C/DEU/CO/6 (2012).
The author has not systematically evaluated this aspect of all country reports by HRC and CERD, or by the Special Rapporteur for Free Speech, and she is not aware of anyone else who has done so.

138. The Special Rapporteur on Free Speech has issued country-monitoring reports that do rigorously evaluate hate speech measures consistent with ICCPR 19(3) and General Comment 34. For example, Special Rapporteur on Free Speech Kaye's 2019 report on Ethiopia concluded that the draft of Ethiopia's Hate Speech and Disinformation Proclamation "goes far beyond the command of [ICCPR] 20(2) and the limitations on restrictions required by [ICCPR] 19(3)," and thus raises concerns that it "will exacerbate ethnic tension, which . . . may fuel further violence." David Kaye, *United Nations Special Rapporteur on the Right to Freedom of Opinion and Expression*, U.N. HUM. RTS. OFF. HIGH COMM'R (Dec. 2019), https://www.ohchr.orhg/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25402&LandID=E.
Kaye also strongly urged a range of non-censorial alternatives as being both consistent with the U.N. treaty provisions and more likely to be effective in countering ethnic tension and violence. *Id.* Moreover, he stressed that the prior Special Rapporteur on Free Speech had reached the same conclusions in the context of multiple other country reports:
[ICCPR] 20(2) provides a specific and internationally-recognized definition of hateful advocacy, which international mechanisms have interpreted. The mandate [i.e., the Special Rapporteur] has found the language of "hate speech," untethered from [ICCPR] 20(2), to be problematic over the course of many years. My predecessor found it common that legal

that fall short. Such proponents must also press Platforms to rigorously align their content moderation policies with General Comment 34 and General Recommendation 35.

E. ICCPR 20(2) requires governments to outlaw only certain designated hate speech

The most important limit on the mandated hate speech restrictions under ICCPR 20(2) and ICERD 4 is the requirement just discussed, that any such restrictions must comply with the strict legitimacy, legality, and necessity requirements under ICCPR 19(3) and elaborated in General Comment 34. It is also noteworthy that, even on its face, ICCPR 20(2) calls for outlawing only a subset of hate speech, which is significantly narrower than the concept of illegal hate speech in many countries' laws.

ICCPR 20(2) provides that, "[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."[139] This language is not as narrow as the U.S. had urged, with the support of many countries from various regions, during the treaty negotiations,[140] but still far narrower than the original Soviet proposal.[141] The U.S. sought to confine the provision to outlawing incitement to violence — a standard that was more speech-protective than what U.S. First Amendment law then required.[142] In

---

frameworks attempting to regulate hate speech risked interfering, and in fact did interfere, with freedom of expression principles, when insufficiently grounded in international human rights law.

Id. (citing 2012 Special Rapporteur on Free Speech, *supra* note 22).

The author of this Article has not systematically reviewed all country-monitoring reports by the Special Rapporteur on Free Speech, nor is she aware of anyone else who has done so.

139. ICCPR, *supra* note 95, 20(2).

140. *See* Cleveland, *supra* note 18, at 224 (noting that the U.S.'s strong free speech position was supported "by many states, including other advanced Western democracies, as well as a number of Latin American states, Lebanon, Japan, and others").

141. *Id.* at 221. Likewise, a Soviet proposal for a broad definition of punishable hate speech in the 1948 Genocide Convention was also rejected in favor of the narrow definition that the U.S. had advocated. *See id.* at 216.

142. Throughout the ICCPR drafting process — from 1945 until 1954 — U.S. free speech law embodied a broader, more malleable concept of punishable incitement than the narrow concept that the U.S. had advocated for the ICCPR. Not until 1969 —three years after the U.N. General Assembly adopted the ICCPR — did the Supreme Court explicitly

contrast, the Soviets sought to criminalize "all racist and nationalist propaganda."[143] Ultimately, the adopted language is significantly more speech-protective than the Soviet proposal in two respects: it rejects both the requirement of criminalization and the broad, vague definition of the targeted expression.

As adopted, ICCPR 20(2) singles out not only "*advocacy* of *hatred*," but also advocacy that "constitutes *incitement* to discrimination, *hostility* or violence."[144] The four italicized words in ICCPR 20(2) have quite limited, specific meanings. This significant point was stressed in an important U.N. guidance, the "Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence" (the "Rabat Plan of Action"). The Rabat Plan of Action was adopted in 2012 under the auspices of the U.N. High Commissioner for Human Rights, as the result of a consultative process involving forty-five experts from different cultural backgrounds and legal traditions.[145] Construing ICCPR 20(2)'s language, the Rabat Plan of Action stated:

> [T]he terms "hatred" and "hostility" refer to intense and irrational emotions of opprobrium, enmity and detestation towards the target group; the term "advocacy" is to be understood as requiring an intention to promote hatred publicly towards the target group; and the term "incitement" refers to statements about national, racial or religious

---

limit this category of punishable speech to expression that intentionally incited imminent violent or lawless conduct that is likely to occur. *See* Brandenburg v. Ohio, 395 U.S. 444 (1969). Before *Brandenburg*, the Court had permitted government to punish speech due to its feared "bad tendency." *See, e.g.*, Dennis v. United States, 341 U.S. 494, 516–17 (1951) (upholding the federal Smith Act, which made it a crime "to knowingly . . . advocate . . . or teach the . . . propriety of overthrowing . . . any government in the United States by Force or violence").

Professor Sarah Cleveland has explained that the U.S. insisted on stronger free speech protection under U.N. law than what existed under U.S. law as a prophylactic measure, because the U.N. "principles would often be applied in states with weaker traditions of free speech and press and judicial independence." Cleveland, *supra* note 18, at 224. This is a noteworthy argument that, far from being less speech-protective than U.S. law, U.N. law should be more speech-protective.

143. Cleveland, *supra* note 18, at 221.

144. ICCPR, *supra* note 95, 20(2) (emphasis added).

145. *Freedom of Expression vs Incitement to Hatred: OHCHR and the Rabat Plan of Action*, U.N. HUM. RTS. OFF. HIGH COMM'R, https://www.ohchr.org.EN/Issues/FreedomOpinion/Articles19-20/Pages/Index.aspx (last visited Feb. 19, 2021).

> groups which create an imminent risk of discrimination, hostility or violence against persons belonging to those groups.[146]

In 2012, then-Special Rapporteur for Free Speech Frank La Rue also stressed the narrow meanings of ICCPR 20(2)'s key terms. Specifically, La Rue concluded, "the following elements [are] essential when determining whether an expression constitutes incitement to hatred: real and imminent danger of violence resulting from the expression; [and the] intent of the speaker to incite discrimination, hostility, or violence."[147] La Rue cautioned, "since not all types of inflammatory, hateful or offensive speech amount to incitement," the broad general concept of hate speech "should not be conflated" with the specific subset of such speech that ICCPR 20(2) requires governments to outlaw.[148]

ICCPR 20(2)'s relatively limited concept of proscribable hate speech can be appreciated by contrasting it with ICERD 4.[149] The latter requires

---

146. Rep. of the U.N. High Comm'r for Hum. Rts. on the Expert Workshops on the Prohibition of Incitement to Nat'l, Racial or Religious Hatred, U.N. Doc. A/HRC/22/17/ADD.4, at 10 n.5 (2013).

147. 2012 Special Rapporteur on Free Speech, *supra* note 22, at ¶ 46. *See also id.* at ¶ 44(e) ("'Hostility' is a manifestation of hatred beyond a mere state of mind"); *id.* ¶ 50(b) ("No one 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."). Incorporating the foregoing narrow definitions of ICCPR 20(2)'s key terms from the Rabat Plan of Action and the Special Rapporteur on Free Speech into ICCPR 20(2)'s language, the result is that ICCPR 20(2) requires governments to outlaw only expression that (1) intentionally promotes intense and irrational emotions of opprobrium, enmity and detestation towards the target national, racial, or religious group, which constitute a manifestation of hatred beyond a mere state of mind, and (2) creates an imminent risk of causing such intense and irrational emotions, or discrimination or violence against persons belonging to those groups.

148. *Id.* ¶ 49; *see also* 2011 Special Rapporteur on Free Speech, *supra* note 12, ¶ 28 (noting that, for speech to be subject to regulation under ICCPR 20(2), there must be a "very close link between the expression and the resulting risk of discrimination, hostility or violence"); *see also* Detailed Guidance, *supra* note 3, at 13 (explaining that ICCPR 20(2) and ICERD 4 require "an intention to promote hatred publicly towards the target group" and "statements which create an imminent risk of discrimination, hostility or violence").

149. The full text of [ICERD] 4 provides:

> States Parties condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive

governments to outlaw not only "incitement to discrimination," as in ICCPR 20(2),[150] but also the "dissemination of ideas based on racial superiority or hatred."[151] On the one hand, ICERD 4's language expressly calls for outlawing a broader category of hate speech than does ICCPR 20(2).[152] However, ICERD 4 also qualifies its entire definition of proscribable hate speech with its "due regard clause." This clause stipulates that any action under ICERD 4 must be undertaken "with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in [ICERD] 5 . . .." The rights expressly set forth in ICERD 5 include "[t]he right to freedom of opinion and expression." Although the U.S. had expressed concern about ICERD 4's potential inconsistency with freedom of speech while it was being negotiated, the U.S. ultimately voted in favor of ICERD 4 specifically

---

> measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in article 5 of this Convention, inter alia:
>
> (a) Shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof;
> (b) Shall 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;
> (c) Shall not permit public authorities or public institutions, national or local, to promote or incite racial discrimination.

ICERD, *supra* note 97.

150. ICCPR, *supra* note 95, 20(2).

151. ICERD, *supra* note 97, art. IV(a).

152. Notwithstanding the apparent breadth of ICERD's "dissemination" clause, General Recommendation 35 indicates that this clause should be construed as narrowly as the "incitement" clause. Gen. Recommendation 35, *supra* note 23, ¶ 16 (explaining that "incitement offenses" should turn on the speaker's "intention" and "the imminent risk or likelihood that the conduct" that the speaker intended "will result from the speech in question," and concluding that these "considerations . . . also apply to the other offences" under ICERD 4, including "dissemination of ideas based on racial or ethnic superiority or hatred."). *See also id.* ¶ 25 (stating that expression "in the context of academic debates, political engagement or similar activity" may not be punished unless it constitutes "incitement to hatred, contempt, violence or discrimination.").

because of its due regard clause.[153] In practice, however, U.N. officials and bodies have not consistently enforced this clause to strictly constrain hate speech restrictions.

It bears repeating that the most important constraint on hate speech restrictions under both ICCPR 20(2) and ICERD 4 is the requirement — under General Comment 34 and General Recommendation 35, respectively — that all such restrictions must comport with ICCPR 19(3), including by constituting the least restrictive alternative to advance a specified important public purpose. In short, *the strict ICCPR 19(3) requirements for any speech restriction are, in effect, incorporated into and strongly qualify the express language of both ICCPR 20(2) and ICERD 4.*

VI. MULTIPLE RECENT REPORTS BY U.N. BODIES AND OFFICIALS HAVE CALLED FOR THE PLATFORMS' CONTENT MODERATION POLICIES (INCLUDING THEIR HATE SPEECH POLICIES) TO BE CONSISTENT WITH SPEECH-PROTECTIVE U.N. STANDARDS

In a 2018 report, Special Rapporteur on Free Speech Kaye urged the Platforms to adhere to U.N. free speech norms, in accordance with the U.N. Guiding Principles on Business and Human Rights (the "Guiding Principles"), which the U.N. Human Rights Council[154] adopted

---

153. *See* Draft Int'l Convention on Elimination of Racial Discrimination, at 152, U.N. Doc. A/C.3/SR.1318 (1965) ("[The U.S. delegate] emphasized that her delegation had been able to support the text only on the understanding that . . . article IV did not impose on a State Party the obligation to take any action impairing the right to freedom of speech and freedom of association.") To underscore this understanding, the U.S. ratification of ICERD was subject to a reservation, specifying that the U.S. "does not accept any obligation . . . to restrict" "individual freedom of speech, expression and association . . . to the extent that they are protected by the Constitution and laws of the United States." Int'l Convention on Elimination of Racial Discrimination, Reports Submitted by State Parties Under Article 9 of the Convention, ¶ 155, U.N. Doc. CERD/C/351/Add.1 (2000).

154. The Council consists of forty-seven Member States that are elected by the majority of members of the U.N. General Assembly. *Membership of Human Rights Council*, U.N. HUM. RTS. COUNCIL, https://www.ohchr.org/en/hrbodies/hrc/pages/membership.aspx (last visited Feb. 19, 2021). The Council's membership is based on a geographical distribution formula, allocating specified numbers of seats among different regions of the world. *Id.* Members serve for three years and are not eligible for immediate re-election after serving two consecutive terms. *Id.*

unanimously in 2011.[155] While acknowledging that the Guiding Principles are non-binding, Kaye explained that the Platforms' "overwhelming role in public life globally argues strongly for . . . [them to] . . . adopt[] and implement[]" these Principles.[156] Stressing the strong protection that ICCPR 19(3) accords to controversial speech, including hate speech, the report maintained that the Platforms should permit "users to develop opinions, express themselves freely and access information of all kinds in a manner consistent with human rights law."[157]

In 2019, Special Rapporteur on Free Speech Kaye issued another report that renewed the call for the Platforms to align their content moderation policies with U.N. law, specifically applying that recommendation to the Platforms' hate speech policies.[158] This report repeatedly stressed the importance of counter speech as an appropriate and effective alternative to speech restrictions. For example, the report stated: "While . . . [online] companies should combat [hateful] attitudes with education, condemnation and other tools, legal restrictions [on hateful expression] will need to meet the strict standards of international human rights law."[159] The report specifically endorsed "as alternatives . . . to the banning of accounts and the removal of content," the following non-censorial measures against hate speech: "tools that promote individual autonomy, security and free expression," which "involve de-amplification, de-monetization, education, counter-speech, reporting and training."[160]

In September 2020, these same points were reiterated by the U.N.'s Detailed Guidance.[161] While the Detailed Guidance applies to hate speech in all contexts, it expressly explains that one important and specific context in which it applies is the Platforms' content moderation policies.

---

155. *Keeping Momentum: One Year in the life of UN Human Rights Council,* HUM. RTS. WATCH (Sept. 22, 2011), https://www.hrw.org/report/2011/09/22/keeping-momentum/one-year-life-un-human-rights-council.

156. 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 5; *accord* Benesch, *supra* note 20, at 90 (recognizing that the Guiding Principles call on companies to "respect" U.N. human rights law, but maintaining that "companies like Facebook should not only respect the law but [also] comply with it, due to their vast regulation of speech, especially public discourse"). *See also* Land 2013, *supra* note 66, at 447 (arguing that the Platforms are directly bound to respect users' free speech rights under ICCPR 19).

157. 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 45.

158. 2019 Special Rapporteur on Free Speech, *supra* note 4 *passim*.

159. *Id.* ¶ 19.

160. *Id.* ¶ 58(f).

161. *See* Aswad 2020, *supra* note 2.

Specifically, the Detailed Guidance stresses that the Platforms may only regulate hate speech consistent with the strict, speech-protective U.N. standards.[162] Noting the necessity/least restrictive alternative prerequisites for suppressing any speech — including hate speech — the Detailed Guidance quotes the key passage from the 2019 Special Rapporteur on Free Speech report, quoted in the preceding paragraph.[163] In short, the Detailed Guidance maintains that the Platforms should not remove hate speech or ban accounts for posting hate speech except as the last resort.

VII. THE PLATFORMS' CONTENT MODERATION POLICIES GOVERNING HATE SPEECH LIKELY VIOLATE U.N. FREE SPEECH STANDARDS

Because the Platforms are constantly revising their content moderation policies, including those regarding hate speech, this Article errs on the side of caution in asserting only that such policies "likely" violate U.N. free speech standards. Moreover, since the Platforms ramped up their restrictions on hate speech only recently, and since the calls for them to align those restrictions with U.N. standards are also quite recent, the process of elaborating on appropriate Platform hate speech policies is likewise at its incipient stage. Encouragingly, civil society organizations and individual experts are increasingly engaging in this crucial process.[164] Even at this early stage, however, there is ample basis to conclude that the Platforms' hate speech policies do not adhere to U.N. standards.

As discussed above, public officials and community members across the political and ideological spectrum complain that Platforms unjustly suppress online expression, particularly in situations in which the violative nature of the suppressed speech is unclear to the Platforms' users.[165] These complaints indicate that the Platforms' content moderation policies run afoul of, at least, the legality requirement of U.N. free speech law as unduly vague formulations that are not narrowly tailored.

Eminent international human rights law experts have undertaken detailed analyses of some Platform hate speech policies, specifying how those policies fall short of the prerequisites for permissible speech

---

162. Detailed Guidance, *supra* note 3, ¶ 68.2 (calling on "tech and social media companies" to "[a]lign their content policies on hate speech with international human rights norms and standards, including the Rabat Plan of Action").

163. *Id.* ¶ 68.6.

164. *See supra* note 21.

165. *See supra* text accompanying notes 51-63.

restrictions under U.N. law. For example, Special Rapporteur on Free Speech Kaye's 2018 report flagged unduly vague policies violative of ICCPR 19(3)'s legality requirement, including Twitter's then-prohibition on speech that "incites fear about a protected group." Kaye found Twitter's policy illustrative of several Platforms' "policies on hate, harassment, and abuse," concluding that such policies "do not clearly indicate what constitutes an offence."[166]

Moreover, in Kaye's 2019 report, which specifically focused on hate speech, he concluded with a more general indictment of the Platforms' hate speech content moderation policies and practices, stating that the Platforms "manage hate speech . . . almost entirely without reference to the human rights implications."[167] Likewise, a 2018 "Policy Brief," issued by the international free speech organization Article 19, analyzed the hate speech restrictions enforced by Facebook, Twitter, YouTube, and Google, and concluded that "their policies overall fall below [the U.N.'s] international standards on freedom of expression."[168]

Sadly, the failure of the Platforms' hate speech policies to comply with ICCPR 19(3)'s legality requirement has an especially adverse impact on the very minority groups that those policies are intended to protect. As Special Rapporteur on Free Speech Kaye's 2018 report concluded: "The vagueness of hate speech and harassment policies has triggered complaints of inconsistent policy enforcement that penalizes minorities while reinforcing the status of dominant or powerful groups."[169]

An especially in-depth assessment of a Platform hate speech policy was undertaken by Professor Aswad.[170] In a 2018 article, Aswad laid out detailed criteria for evaluating whether any social media company's hate speech policy complies with U.N. free speech law. For example, to comply with the necessity requirement, she "proposes that a company would need . . . to publicly commit to three steps."[171] Significantly, this three-step

166. 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 26.

167. 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 42.

168. ARTICLE 19, *supra* note 11, at 18. *See also* Benesch, *supra* note 20, at 103 (concluding that "most platforms' [content moderation] rules manifestly do not meet" the legality requirement).

169. 2018 Special Rapporteur on Free Speech, *supra* note 12, ¶ 26.

170. Aswad holds the Herman G. Kaiser Chair in International Law at the University of Oklahoma's College of Law, and previously served as the director of the human rights law office at the U.S. Department of State.

171. Aswad 2018, *supra* note 13, at 47.

analysis for assessing whether a Platform's content moderation policy complies with the necessity requirement under U.N. free speech law was explicitly endorsed in Special Rapporteur on Free Speech Kaye's 2019 report:[172]

> Evelyn Aswad identifies three steps that a company should take under the necessity framework: [1] evaluate the tools it has available to protect a legitimate objective without interfering with the speech itself; [2] identify the tool that least intrudes on speech; and [3] assess whether and demonstrate that the measure it selects actually achieves its goals. This kind of evaluation is in line with the call made in the Guiding Principles on Business and Human Rights for businesses to ensure that they prevent or mitigate harms.[173]

Elaborating on the first element of the necessity assessment, Professor Aswad explained that, as an alternative to restricting speech, Platforms "can . . . speak out on issues, educate users, and promote dialogue."[174] Turning to the second element, Professor Aswad explained that, if these non-censorial alternatives proved ineffective so that the Platform would have to resort to some speech suppression, then it "should carefully develop a continuum of options . . . and commit publicly to selecting the least intrusive" among them:

> For example, a company could give its users a means to opt out of offensive material. [Or] a company [could] avoid[] giving problematic posts a circulation boost, but . . . not delete them or affect its users' ability to circulate the posts. A company could also lower the ranking of problematic posts in search results or otherwise decrease their visibility . . . Where speech must be banned, geo-blocking . . . a particular post from view in the particular country could be considered (rather than removing the information from the platform). A more intrusive infringement . . . would be to delete a post but to allow the speaker to continue to speak on the platform. Warnings could be issued to a user who repeatedly violates a company's speech code before taking more

172. 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 52.
173. *Id.*
174. Aswad 2018, *supra* note 13, at 48.

> severe measures. The most extreme end of the continuum may be blocking a user's account in egregious situations.[175]

Finally, concerning the third element, Aswad explains that a Platform should "diligently monitor[] whether the measure it has selected is" actually effective, without "negative unintended consequences that may outweigh the desired benefits."[176]

Applying the necessity requirement to Twitter's hate speech policy through such a rigorous analysis, Professor Aswad concludes that the policy fails.[177] Undertaking a similarly probing analysis of Twitter's hate speech policy under U.N. free speech law's legality requirement, Professor Aswad concludes that the policy likely also fails this element of ICCPR 19(3)'s three-part test.[178]

A 2020 article by Professors Molly Land and Rebecca Hamilton, aptly entitled *Beyond Takedown*, well complements Professor Aswad's analysis of U.N. free speech law's necessity/least restrictive alternative requirement, as endorsed by Special Rapporteur on Free Speech Kaye. Land and Hamilton lay out many specific non-censorial components of the Platforms' potential "toolkit for responding to online hate," which "involve a broader range of actors [beyond the Platforms themselves], and which draw on a mix of online and offline methods both to prevent hate speech before it occurs and to minimize its impact after it occurs."[179] Their intriguing suggestions, drawing on actual initiatives that have been implemented by Platforms and civil society organizations, are grouped into three categories: the Platforms' technical design features; user education; and counterspeech. For example, Land and Hamilton suggest that the Platforms should explore the following technical design features:

> Platforms could be designed in ways that work to minimize the online disinhibition effect, such as through the use of cues reminding users of their shared humanity . . . Algorithms could be tweaked to deprioritize particularly extreme or virulent content . . . Warnings or other kinds of prompts that question the user as to whether they really want to post

175. *Id.* at 49–50.
176. *Id.* at 51.
177. *Id.* at 52.
178. *Id.* at 46–47.
179. Molly Land & Rebecca Hamilton, *Beyond Takedown: Expanding the Toolkit for Responding to Online Hate*, 143 PROPAGANDA, WAR CRIMES TRIALS & INT'L LAW: FROM COGNITION TO CRIMINALITY 1, 8 (2020).

certain content may be enough to deter at least some of those who would otherwise post hate speech.[180]

VIII. CONCLUSION

As U.N. Special Rapporteur on Free Speech David Kaye concluded in his 2019 report, the Platforms "have for too long avoided [U.N.] human rights law as a guide to their rules . . ., notwithstanding the extensive impacts they have on the human rights of their users and the public."[181] Since the Platforms are private entities, their power to restrict users' speech is not constrained by the First Amendment free speech guarantee, the Fifth Amendment due process guarantee, the Fourteenth Amendment equal protection guarantee, or any other limit that the U.S. Constitution imposes on government action. Accordingly, the Platforms may wield their unprecedented censorial power in a manner that is substantively arbitrary and discriminatory, and procedurally unfair. Indeed, a steady stream of evidence demonstrates that they are in fact doing so.

Given the Platforms' preeminence as venues for free speech and for the political discourse that is the lifeblood of democratic self-government, their unprecedented and unconstrained censorial powers threaten essential foundations of liberal democracies: free speech, equality, due process of law, and the rule of law. It is crucial to pursue an alternative means to rein in that power. A promising approach, consistent with the U.N. Guiding Principles on Business and Human Rights, is for the Platforms to honor and advance the key speech-protective aspects of U.N. free speech law. Along with U.S. law, these important elements of U.N. law reflect a view shared by many human rights advocates around the world: that the most constructive way to combat hateful attitudes and action is through robust, proactive, non-censorial approaches, supplemented by narrowly limited speech restrictions.

Free speech proponents must continue to promote the speech-protective elements of the U.N. free speech regime, urging that those elements be more clearly and consistently implemented throughout the U.N. system, and insisting on the Platforms' adherence to them.

---

180. *Id.* at 9.

181. 2019 Special Rapporteur on Free Speech, *supra* note 4, ¶ 58.

MICHIGAN STATE UNIVERSITY
COLLEGE OF LAW

INTERNATIONAL
LAW REVIEW

Cite as MICH. ST. INT'L L. REV.

The Michigan State International Law Review (ISSN 2328-2991 (print); ISSN 2328-3068 (online)) is published three times a year by the students of the Michigan State University College of Law. The Editorial and Business offices are located at:

Michigan State University International Law Review
648 N. Shaw Lane, Room 209
East Lansing, MI 48824-1300
Phone: 517-432-6932
E-mail: ilr@msu.edu
Webpage: www.msuilr.org

Subscriptions: The Michigan State International Law Review ("International Law Review") is a digital publication. Individual articles are available online at http://digitalcommons.law.msu.edu/ilr. Print copies may be ordered directly from the journal.

Single Issues: Single issues of current and future volumes may be purchased by contacting the journal directly. Queries regarding back issues of the journal should be directed to the Institutional Repository Officer, 648 N. Shaw Lane, Room 115B, East Lansing, MI 48824.

Production: Footnotes in the International Law Review generally conform to The BlueBook: A Uniform System of Citation (21st ed. 2020). The International Law Review is published on Digital Commons, and print copies are produced by Lulu.com.

Disclaimer: Opinions expressed in essays, articles and comments are solely those of the respective authors. The opinions are not presented as the views of this publication, its editors or the Michigan State University College of Law. The journal welcomes comments on this and every volume.

MICHIGAN STATE UNIVERSITY
COLLEGE OF LAW

BOARD OF TRUSTEES

President of Michigan State University and MSU College of Law
Donald Nystrom, '00, Chair
Charles A. Janssen, Vice-Chair
Elaine Fieldman, '76
Clifton E. Haley, '61, President Emeritus
Hon. Wallace B. Jefferson
Maurice G. Jenkins, '81
Charles E. Langton, '87
John J. Latella, '00
Douglas Laycock
Cary S. McGehee, '89
Hon. David W. McKeague
Richard D. McLellan
Hon. Collen M. McNamara
Mayer (Mike) Morganroth, '54
Michael G. Morris, '81
Bryan T. Newland, '07
Linda M. Orlans, '87, President Emerita
John D. Pirich
Hon. David L. Porteous
Jennifer Poteat, '04
Hon. G. Scott Romney
Robert Worthington, '07

TRUSTEES EMERITI

Hon. Marianne O. Battani, '72
Raymond R. Behan, '60
Joseph J. Buttigieg III, '75
Frederick D. Dilley, '75
Edwin W. Jakeway, '61
Hon. Norman L. Lippitt, '60
John D. O'Hair, '54
Peter J. Palmer, '68
Kenneth J. Robinson
John F. Schaefer, '69
Hon. Richard F. Suhrheinrich, '63, President Emeritus

LAW COLLEGE ADMINISTRATION

June P. Youatt, Provost of Michigan State University and MSU College of Law
Lawrence Ponoroff, Dean and Professor of Law
Daniel Barnhizer, Associate Dean for Graduate and International Programs, Professor of Law, and the Bradford Stone Faculty Scholar
Melanie B. Jacobs, Senior Associate Dean for Academic Affairs and Professor of Law
Carrie Feehely, Assistant Dean of Admissions and International Programs
Mary A. Ferguson, Assistant Dean for Diversity and Equity Services, Title IX Coordinator
Richard Lameti, Associate Dean of Finance and Administration
Michael Sant'Ambrogio, Associate Dean of Research and Associate Professor of Law
Abijah P. Taylor, Associate Dean for Student and Academic Affairs
Maria Stein, Assistant Dean for Career Development

FACULTY

Daniel D. Barnhizer
Bruce W. Bean
Mary A. Bedikian
Susan H. Bitensky
David Blankfein-Tabachnick
Kristine L. Bowman

Adam Candeub
Jennifer Carter-Johnson
James M. Chen
Jennifer L. Copland
Nancy A. Costello
Tiffani Darden
David S. Favre
Matthew L.M. Fletcher
Jeremy Francis
Brian Gilmore
Catherine M. Grosso
Melanie B. Jacobs
Clark C. Johnson
Brian C. Kalt
Joseph Kozakiewicz
Mae Kuykendall
Stephanie J. LaRose
Deanne Andrews Lawrence
Michael A. Lawrence
Anne Lawton
Daniel E. Manville
Amy C. McCormick
Nicholas Mercuro
Geeyoung Min
Noga Morag-Levine
Carney Anne Nasser
Barbara O'Brien
Daphne O'Regan
Sean A. Pager
Kathleen E. Payne
Lawrence Ponoroff
Philip A. Pucillo
Frank S. Ravitch
John W. Reifenberg, Jr.
Carla L. Reyes
Jennifer Rosa
Michael Sant'Ambrogio
Kevin W. Saunders
Elliot A. Spoon
Cynthia Lee Starnes
Glen Staszewski
Paul Stokstad
Charles J. Ten Brink
David B. Thronson
Veronica Thronson
Christina Wease
Joshua Wease
Nicholas J. Wittner

PROFESSORS EMERITI

Clark C. Johnson
Kevin C. Kennedy
Amy C. McCormick
Robert A. McCormick
Matthew McKinnon
Kathleen E. Payne
Brenda Jones Quick

ADJUNCT FACULTY

Rosemarie Aquilina
Joseph C. Basta
Michael Behan
Drew Bowden
Erica G. Brown
Caroline Bruce-Erickson
Cynthia C. Bullington
Paul Burakoff
Juan Caballero
Melissa Card-Abela
Jeff Carter-Johnson
Bradley N. Deacon
Ronald Deneweth
Todd Dishman
Brent Domann
Allison Eicher
Adam Ekonomon
Trent English
Ezekiel Fletcher
Kathryn E. Fort
Neal D. Fortin
Kevin S. Gentry
Andrew Gerdes
Elizabeth G. Goebel
P. Greg Gulick
Curtis Hall
Hildur Hanna
P. Vincent Hegarty
Glenford Jameson
David Jukes
Steven Kaplan
Dennis Kennedy
Mohammad Khalil
Jesse Kirchner

Joanna Kloet
Richard Lameti
Jonathan E. Lauderbach
David McKeague
Veronica Valentine McNally
Jane Meland
Linda Mettes
Brooke Moynihan Minore
Umar Moghul
Mikhail Murshak
Brian A. Pappas
John Pirich
Antoinette Raheem
David Revore
Deborah M. Schneider
Lawrence P. Schweitzer
Mary Kay Scullion
Ann Sherman
Stephen H. Sinas
George T. Sinas
Elaine Spiliopoulos
Brock Swartzle
Daryl Thompson
Melanie VanAntwerp
Bryan J. Waldman
J. Dallas Winegarden

ALUMNI ASSOCIATION

For over 100 years, Michigan State University College of Law, formerly known as Detroit College of Law, has provided a legal education which prepares qualified students not only for the Bar, but for a lifelong career. Today, the College continues this tradition of excellence which is exemplified by the achievement of over 9,000 alumni in private practice, on the bench, in the classroom, and in the corporate world.

Providing this education is, of course, difficult. Even more difficult, however, is the task of assuring that an excellent legal education is available at a cost that will attract bright, deserving students, regardless of their gender, race, or economic background. This proud tradition of equal opportunity is, in fact, as old as the school itself, dating back to the founding of the Detroit College of Law in 1891.

Helping to preserve the College's legacy of educational excellence and equal opportunity are two important volunteer groups: the Alumni Association and the Office of Advancement. Together they provide continuing financial assistance and volunteer leadership. Their efforts provide the means to bridge the crucial gap between operating expenses and revenues. Their financial support and management expertise enable MSU College of Law to keep tuition costs down, to preserve and enhance the quality of our faculty and our facility, to promote professional and personal fellowship among alumni, and to help provide scholarship assistance to those students who merit such aid.

Your interest in the Alumni Association is greatly appreciated. For more information, contact the Office of Advancement of Michigan State University College of Law at (517) 432-6840.

ALUMNI ASSOCIATION BOARD 2019-2020

Michael Daum, President, '11
Christopher Jackson, Vice President, '13
Daniel Elliott, Treasurer '16
Robert Kirk, Director of Fundraising '83
Christopher Giles, Director of Alumni Engagement '02
Victoria Convertino, Secretary, '16
Mallory Field, Parliamentarian, '11

ALUMNI ASSOCIATION PARLIAMENTARIAN BOARD

Liam Burke, '07
John Conniff, '11
Chantel Flegler, '16
Courtney Gabbara, '12
Denise Hirschmann, '92
Christopher Johnson, '11
Fedor Kozlov, '10
Kate Matych, '16
René Roupinian, '94
Alex Rusek, '13
Robert Shapiro, '94
Brandon Sherman, '04
Jacob Simon, '17
Stephanie Vicent, '11

Ex Officio:
Dean Lawrence Ponoroff
Lauren Wilton

MICHIGAN STATE UNIVERSITY COLLEGE OF LAW

INTERNATIONAL LAW REVIEW

MASTHEAD
2020-2021