Washington and Lee Law Review

Volume 77 | Issue 2 Article 4

4-27-2020

To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?

Evelyn Mary Aswad
University of Oklahoma College of Law

To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?

Evelyn Mary Aswad*

Abstract

Global social media platforms are grappling with whether to align their corporate speech codes with international human rights law. Facebook's June 2019 report that summarized worldwide feedback about its proposed independent oversight board for content moderation noted a split in stakeholder opinions on this topic. The UN's top expert on freedom of expression as well as many civil society members recommended that Facebook anchor its content moderation in the international human rights law regime. Others expressed concern that this legal regime would not be sufficiently protective of speech and contained inconsistencies that create problems for content moderation.

Those concerns were linked to a recent scholarly call for updates to the UN's international legal regime regarding

* The author is the Herman G. Kaiser Chair in International Law and the Director of the Center for International Business & Human Rights at the University of Oklahoma's College of Law. Previously, she was the director of the human rights law office at the U.S. Department of State (2010–2013) and served as an attorney-adviser in that human rights law office from 2004–2009. The author wishes to thank Stanford University's Global Digital Policy Incubator for inviting her to participate in a variety of events in which issues at the intersection of content moderation and international human rights law were discussed, which helped with her analysis of the issues in this Article. The author also wishes to thank her research assistant, Bailey Betz, for his excellent assistance with this Article. The views are solely those of the author.

freedom of expression, particularly with respect to the International Covenant on Civil and Political Rights and the Convention on the Elimination of All Forms of Racial Discrimination. This Article examines the scholarly call's analysis to assess whether its conclusions are correct, which would make this body of law less useful for platforms to adopt in content moderation.

This Article finds that the state of international law on freedom of expression is more protective of speech (and more coherent) than the scholars assessed and proposes ways to achieve their laudable goal of promoting broad protections for freedom of expression in international law. The Article concludes that the existing international legal regime on freedom of expression remains a useful resource for content moderation by global platforms.

Table of Contents

I. Introduction

In the midst of controversy regarding how global social media platforms curate user-generated content,[1] these companies are considering whether they should align their speech codes for online content with international law standards on expression. For example, in June 2019, Facebook released a report that summarized worldwide feedback on its proposed independent oversight board, which will review some of the platform's content moderation decisions.[2] The Facebook

1. *See, e.g.*, Cat Zakrzewski, *The Technology 202: Lawmakers Plan to Ratchet Up Pressure on Tech Companies' Content Moderation Practices*, WASH. POST (Apr. 9, 2019), https://perma.cc/K7CX-BHXX (discussing a variety of criticisms of platform content moderation with respect to potentially harmful and offensive speech) (on file with the Washington and Lee Law Review).

2. *See* FACEBOOK, GLOBAL FEEDBACK & INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS 10–37 (2019), https://perma.cc/2W84-6XAX (PDF) [hereinafter FACEBOOK REPORT] (summarizing views of external stakeholders on a variety of aspects of the proposed oversight board). In November 2018, Facebook CEO Mark Zuckerberg proposed the establishment of an *independent* body that would assess speech issues on the company's platform. *Id.* at 4. Harvard Law School Professor Noah Feldman had initially proposed the idea of a Facebook "Supreme Court" to ensure an "independent decision-making process based on a formal commitment to freedom of expression" to adjudicate speech issues arising from content generated by the billions of individuals who use the platform. *Id.* at 8. In January 2019, Facebook issued a proposed charter for the new body (the "Oversight Board") and conducted six months of consultations throughout the world to receive feedback. *Id.* at 4–5. The

report notes that the United Nations' (UN) top expert for freedom of expression and members of civil society,[3] including myself,[4] advocated for use of international human rights law's free expression protections in the Oversight Board's review of the company's global content moderation decisions.[5] These stakeholders generally linked their views to the UN Guiding Principles for Business & Human Rights[6] (UNGPs), a framework that sets forth the international community's expectations for corporations when their operations intersect with human rights issues.[7] Under the UNGPs, companies are

---

Oversight Board's mandate will cover content decisions made by the company under its corporate speech code (i.e., the Community Standards) and not content removals based on compliance with local laws. *Id.* at 35.

3. *See id*. at 34 (discussing external stakeholder support for Facebook to adopt UN human rights standards in content moderation). Civil society groups have also called for the creation of "social media councils" that would provide independent oversight over the content moderation of all platforms. *See id.* at 8 ("[An NGO called] Article 19 has proposed establishing a multi-stakeholder and industry-wide "Social Media Council," which shares some similarities with the Oversight Board proposed by Facebook."); STANFORD GLOBAL DIGITAL POLICY INCUBATOR, ARTICLE 19, & UNITED NATIONS SPECIAL RAPPORTEUR ON FREEDOM OF OPINION AND EXPRESSION, SOCIAL MEDIA COUNCILS: FROM CONCEPT TO REALITY 10 (2019), https://perma.cc/X2CG-QZYQ (PDF) (endorsing the creation of social media councils grounded in international human rights law principles for freedom of expression). On September 17, 2019, Facebook released the charter for its Oversight Board, which states that "the board will pay particular attention to the impact of removing content in light of human rights norms protecting free expression." FACEBOOK, OVERSIGHT BOARD CHARTER 5 (2019), https://perma.cc/5J3L-THD7 (PDF). It remains to be seen how Facebook's Oversight Board will interpret this mandate and if other platforms will seek to align their content moderation with international human rights standards.

4. *See* FACEBOOK REPORT, *supra* note 2, at 43 nn. 250 & 257 (citing Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26 (2018)).

5. *See supra* notes 3–4 and accompanying text.

6. John Ruggie, Rep. of the Special Representative of the Sec'y-Gen. on the Issue of Human Rights and Transnational Corps. and Other Bus. Enters., U.N. Doc. A/HRC/17/31 (Mar. 21, 2011) [hereinafter UNGPs].

7. *See* Human Rights Council Res. 17/4, U.N. Doc. A/HRC/RES/17/4, at 2 (July 6, 2011) (endorsing "the Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect, and Remedy' Framework"); *see generally* UNGPs, *supra* note 6 (introducing and explaining the UN Guiding Principles).

expected to "respect" the human rights in key UN instruments[8] in their business operations, which means they "should avoid infringing on the human rights of others" and "should address adverse human rights impacts with which they are involved."[9]

Facebook's report, however, noted that other stakeholders raised concerns about anchoring content moderation to international human rights law.[10] For example, one stakeholder expressed concerns that human rights law could affect how Facebook "balances users' security and trust against freedom of expression."[11] The concerns included whether international human rights law might be too protective of speech and thus limit Facebook's ability to remove speech that offends individuals and advertisers who use the platform.[12] On

8. Principle 12 of the UNGPs calls on business enterprises to respect "international human rights" that are defined to include the International Bill of Human Rights (i.e., the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the International Covenant on Economic, Social, and Cultural Rights) as well as the principles concerning fundamental rights set out in the International Labor Organization's Declaration on Fundamental Principles and Rights at Work, UNGPs, *supra* note 6, at 13. The official commentary to Principle 12 states that businesses may also need to refer to additional UN human rights instruments. *Id*. at 13–14.

9. *Id.* at 13. The U.S. government has endorsed the UNGPs and encouraged their implementation by American companies. U.S. Dep't of State, Bureau of Democracy, H. R. and Lab., U.S. Government Approach on Business and Human Rights 4 (2013), https://perma.cc/BTN8-YWEY (stating the U.S. government's position that the UNGPs should serve "as a 'floor' rather than a 'ceiling' for addressing issues of business and human rights") (on file with the Washington and Lee Law Review); U.S. GOVERNMENT, RESPONSIBLE BUSINESS CONDUCT: FIRST NATIONAL ACTION PLAN FOR THE UNITED STATES OF AMERICA 17 (2016), https://perma.cc/Q95N-FQ8Q (PDF) (encouraging businesses to use "the UN Guiding Principles as a floor rather than a ceiling for implementing responsible business practices").

10. *See* FACEBOOK REPORT, *supra* note 2, at 34 (noting that other stakeholders "felt that focusing exclusively on a human rights framework would not be enough to adjudicate hard cases").

11. *Id.*

12. *See* FACEBOOK, GLOBAL FEEDBACK AND INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS: APPENDIX 123–24 (2019), https://perma.cc/6YT9-UGSW (PDF) (noting concerns of external stakeholders that adherence to international free expression standards would require Facebook to protect speech that it otherwise would remove from its platform under existing rules and policies).

the other hand, another stakeholder cautioned that social media companies may not be well-served by anchoring their private speech codes to existing international human rights law principles because of concerns that this body of law contains conflicting norms and may insufficiently protect speech.[13] In articulating these concerns, this stakeholder cited a 2017 article titled *The Right to Insult in International Law*.[14]

In that article, the authors eloquently present their vision for broad protections for freedom of expression under international law and call for revisions to the regime.[15] After an excellent survey of domestic laws from around the world that fail to protect the "right to insult" rulers, religion, or royalty,[16] the article examines international standards and contends that they not only fail to afford insults sufficient protection, but also provide inconsistent guidance on

13. *See* FACEBOOK REPORT, *supra* note 2, at 34–35 (stating that others took the position that the international "human rights law [regime] is not 'a single, self-contained and cohesive body of rules.' . . . '[T]hese laws are found in a variety of international and regional treaties that are subject to differing interpretations by states that are parties to the convention as well as international tribunals that apply the laws." (quoting Evelyn Douek, *U.N. Special Rapporteur's Latest Report on Online Content Regulation Calls for 'Human Rights by Default*,' LAWFARE BLOG (June 6, 2018, 8:00 AM), https://perma.cc/2FZU-4H5T (last visited Mar. 18, 2020) (noting as well that some commentators are concerned that international freedom of expression standards may not protect speech enough) (on file with the Washington and Lee Law Review).

14. Amal Clooney & Philippa Webb, *The Right to Insult in International Law*, 48 COLUM. HUM. RTS. L. REV. 1 (2017).

15. *See id.* at 47–51 (calling for a revised international law approach to speech that is inspired by the U.S. Constitution's First Amendment); *see also* Alan Wehbé, *Increasing International Legal Protection for Freedom of Expression*, 8 NOTRE DAME J. INT'L & COMP. L. 44, 59 (2018) (arguing for a new treaty on freedom of expression because of concerns that the current regime is not sufficiently protective of speech and not effective in changing problematic governmental approaches to speech).

16. *See* Clooney & Webb, *supra* note 14, at 3–13 (discussing domestic laws that use criminal defamation, sedition, hate speech, "fake" news, and public order as a justification to quash criticism of rulers; criminal blasphemy as well as laws protecting religious feelings to prevent insults to religions; and criminal bans on exhibiting disrespect for royalty).

permissible restrictions for such offensive speech.[17] To achieve broader speech protections, the authors recommend several approaches at the international level.[18] For example, they call on states to seek amendments to existing treaty text or enter reservations to particular treaty provisions,[19] with the ultimate goal of nudging international law closer to U.S. First Amendment protections for speech.[20] The authors conclude by recommending that social media companies respect the international human rights law regime in assessing speech issues on their platforms, but that they do so by incorporating the authors' proposed updates to the regime.[21]

While I share the authors' laudable goal of promoting the broadest possible international protections for freedom of expression, I write this Article to advance the proposition that the state of international law on the ability of individuals to insult or otherwise engage in offensive speech is not bleak, as portrayed in their article, but rather this body of law contains robust and principled speech protections.[22] This Article also notes that a number of the authors' recommended solutions may not be necessary or feasible, proposes more fruitful ways forward in achieving their laudable goal, and defends the

17. *See id*. at 37 ("[T]he right to insult is not sufficiently protected under international law . . . . [I]nternational standards have proved to be confusing . . . .").

18. *See infra* notes 19–21 and accompanying text.

19. *See* Clooney & Webb, *supra* note 14, at 53–54 (calling for deletion of Article 4 of the Convention on the Elimination of Racial Discrimination and for states to opt out of Article 20 of the International Covenant on Civil and Political Rights).

20. *See id*. at 51 ("We believe that international law on insulting speech should be applied in a manner that is . . . more protective of speech, in line with the approach espoused by the U.S. Supreme Court under the First Amendment.").

21. *See id.* at 54 (advocating that their proposed revisions to the international human rights law regime on expression "should also guide the private sector when they are requested to remove online content because it is considered insulting by a person, institution, or government").

22. *Compare infra* Part II, *with* Clooney & Webb, *supra* note 14, at 37–51 (describing the current speech regime in international law and criticizing its inefficacy).

utility of the existing international legal regime in content moderation by private platforms.[23]

Part II of this Article explains why the current state of international law affords more robust protections for insulting speech than Clooney and Webb claim as well as why international standards are generally consistent on this topic. Part III evaluates the authors' recommendations for action at the international level and finds they may not reflect the best way to achieve the authors' important goal. Part III also proposes that the most fruitful path for protecting the ability of individuals to engage in insulting or otherwise offensive speech (and, more generally, to promote broad free expression protections) involves seizing victory from the jaws of defeat by invoking and promoting existing UN treaty texts and positive interpretations from the UN human rights machinery on freedom of expression. Part IV concludes by reaffirming the utility of this body of international law as a key resource in global corporate content moderation.

II. The Current State of International Law on Insults or Otherwise Offensive Speech

The authors of *The Right to Insult in International Law* highlight that the key international human rights treaty provisions involving freedom of expression are Articles 19 and 20 of the International Covenant on Civil and Political Rights[24] (ICCPR) as well as Article 4 of the Convention on Elimination of All Forms of Racial Discrimination[25] (CERD).[26] The authors conclude that these provisions provide insufficient protection

23. *See infra* Parts II, III, IV.

24. International Covenant on Civil and Political Rights, *opened for signature* Dec. 16, 1966, arts. 19, 20, S. Exec. Doc. E, 95-2 (1978), 999 U.N.T.S. 171, 178 (entered into force Mar. 23, 1976) [hereinafter ICCPR].

25. International Convention on the Elimination of All Forms of Racial Discrimination, *opened for signature* Mar. 7, 1966, art. 4, S. Exec. Doc. C, 95-2 (1978), 660 U.N.T.S. 195, 218 (entered into force Jan. 4, 1969) [hereinafter CERD].

26. *See* Clooney & Webb, *supra* note 14, at 15–19 (discussing international law's protections for expression as well as its mandatory bans on hateful speech).

for insulting speech[27] and that the interpretative guidance with respect to these provisions is conflicting.[28] However, their overview of these treaty provisions did not include an analysis of ICCPR Article 19's rigorous tripartite test for speech restrictions or a number of key interpretations issued by the UN human rights bodies charged with monitoring implementation of those treaties.[29] ICCPR Article 19's tripartite test, buttressed by the UN machinery's recommended interpretations, places a substantial legal limitation on the discretion of states to restrict speech, particularly on the topic of insults to rulers, religion, and royalty.[30]

Part II.A discusses how ICCPR Article 19's tripartite test helps achieve the authors' goal of promoting broad protections for expression, including insulting or otherwise offensive speech. Part II.B explains why it is inappropriate to conflate interpretations of the UN's *international* human rights machinery with those issued by *regional* bodies in order to conclude that the UN's *international* law standards are inconsistent. In addition, Part II.B demonstrates that the cited inconsistences within the UN watchdogs' interpretations of freedom of expression are more suitably attributed to an evolution in the thinking of these bodies over time rather than unworkable inconsistencies.

---

27. *See, e.g.*, *id*. at 16 (asserting that the relevant human rights law treaty provisions "allow insulting speech to be silenced if it constitutes discriminatory 'hate speech' targeting minorities on the basis of race, nationality or religion even if there is no criminal intent or risk that it will lead to violence")*; id.* at 38 (concluding that ICCPR Article 20(2) and CERD Article 4 mandate overly broad restrictions on speech).

28. *See id*. at 44–47 (outlining the inconsistencies between the UN human rights machinery and regional human rights mechanisms).

29. *See id*. at 15–21 (discussing the history and a textual analysis of ICCPR Articles 19 and 20 as well as CERD Article 4). Similarly, another scholar who recently called for updates to the international legal regime on freedom of expression concluded that the existing regime was insufficiently protective of speech without analyzing or discussing ICCPR Article 19's tripartite test for speech restrictions. Wehbé, *supra* note 15, at 50–53.

30. *See* Evelyn M. Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 35–37 (2018) (introducing, defining, and discussing the mechanics of the ICCPR Article 19 tripartite test).

A. ICCPR Article 19's Tripartite Test and Its Impact on Bans on Insulting or Offensive Speech

ICCPR Article 19(2) provides broad protections for speech by declaring an expansive right to seek and receive information of all kinds, regardless of frontiers and through any media.[31] However, the article's next clause permits, but does not require, State Parties to limit speech when each part of a tripartite test is met.[32] To be valid, speech restrictions must be (1) "provided by law" and (2) "necessary" to (3) achieve one of the treaty's stated public interest objectives.[33] These three prongs are often referred to as the "cumulative conditions of legality, necessity, and legitimacy."[34]

In 2011, the UN Human Rights Committee, the body of independent experts elected by ICCPR State Parties to monitor implementation of the treaty, recommended interpretations of this article that broadly protect speech rights.[35] According to the Committee, a government must prove that any limitation on expression, including insulting speech, meets Article 19's tripartite test.[36] The remainder of

---

31. *See* ICCPR, *supra note* 24, at art. 19(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.").

32. *See id.* art. 19(3).

33. *See id*. (listing the public interest objectives as the protection of the rights or reputations of others, national security, public order, public health or morals).

34. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 8, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018) [hereinafter *UN SR 2018 Report*].

35. *See* U.N. Human Rights Comm., General Comment No. 34, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011) [hereinafter GC 34] (discussing the scope of freedom of expression and forbidding the adoption of overly broad speech restrictions).

36. *See id.* ¶ 27 ("It is for the State party to demonstrate the legal basis for any restrictions imposed on freedom of expression. If, with regard to a particular State party, the Committee has to consider whether a particular restriction is imposed by law, the State party should provide details of the law and of actions that fall within the scope of the law."); *see also id.* ¶ 35

Part II.A elaborates on Article 19(3)'s tripartite test, highlights key interpretations by the UN human rights machinery, and explains how these three conditions promote broad protections for expression that are in line with the authors' vision of protecting insulting speech.

1. The "Legality" Condition: A Prohibition on Vague Content Bans

The first prong of the tripartite test, the "provided by law" or "legality" condition, is generally understood to mean, *inter alia*, that laws limiting speech must be properly promulgated and not vague.[37] The UN Human Rights Committee has emphasized that legal restrictions on speech must give individuals sufficient notice of the parameters of unacceptable expression.[38] The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (the "UN Special Rapporteur"), an independent expert appointed by the UN Human Rights Council to examine and report on this right, has also emphasized this point.[39] In this

---

> When a State party invokes a legitimate ground for restriction of freedom of expression, it must 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.

37. *See infra* notes 38–41 and accompanying text.

38. *See* GC 34, *supra* note 35, at ¶ 25

> [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 and it must be made accessible to the public. A law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution. Laws must provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not.

Similarly, the United States government interprets the legality condition as encompassing "laws that are accessible, clear, and subject to judicial scrutiny," OFFICE OF THE LEGAL ADVISER, U.S. DEP'T OF STATE, DIGEST OF UNITED STATES PRACTICE IN INTERNATIONAL LAW 226–27 (2011) [hereinafter U.S. OBSERVATIONS].

39. *See UN SR 2018 Report*, *supra* note 34, at ¶ 7 (noting the legality condition means legal restrictions "must be adopted by regular legal

regard, he has criticized many national laws as unduly vague, including those that ban "obscene, gory or offensive material which is likely to cause fear and alarm to the general public" or the "glorification of terrorism."[40] In his report recommending that social media companies align their speech codes with ICCPR Article 19, the UN Special Rapporteur also criticized these private sector codes for incorporating a variety of vague bans on insulting or offensive speech.[41]

As a former President of the American Civil Liberties Union discusses in her recent book on hate speech, prohibiting vague wording in speech bans plays an important role in limiting governmental regulation of offensive speech such as insults.[42] In examining U.S. Supreme Court jurisprudence, she recounts that the Court has repeatedly struck down laws banning "insulting," "abusive," and other such language on

---

processes and limit government discretion in a manner that distinguishes between lawful and unlawful expression with 'sufficient precision'").

40. David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 14, U.N. Doc. A/HRC/71/373 (Sept. 6, 2016) [hereinafter *UN SR 2016 Report*]. The UN Human Rights Committee has also expressed concern about vague phrasing in counter-terrorism laws. *See* GC 34, *supra* note 35, at ¶ 46 ("Such offences as 'encouragement of terrorism' and 'extremist activity' as well as offences of 'praising', 'glorifying', or 'justifying' terrorism, should be clearly defined to ensure that they do not lead to unnecessary or disproportionate interference with freedom of expression."). The Special Rapporteur has also criticized Europe's regional human rights law as failing to define hate speech with sufficient particularity. *See UN SR 2016 Report*, at ¶ 25.

41. *See UN SR 2018 Report*, *supra* note 34, at ¶ 26

> Company prohibitions of threatening or promoting terrorism, supporting or praising leaders of dangerous organizations and content that promotes terrorist acts or incites violence are, like counter-terrorism legislation, excessively vague. Company policies on hate, harassment and abuse also do not clearly indicate what constitutes an offence. Twitter's prohibition of "behavior that incites fear about a protected group" and Facebook's distinction between "direct attacks" on protected characteristics and merely "distasteful or offensive content" are subjective and unstable bases for content moderation.

42. *See* NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 69–72 (2018) (explaining the role of the vagueness standard in U.S. First Amendment jurisprudence).

grounds of ambiguity[43] and assesses that bans on such speech are "insusceptible to precise, narrow definition."[44] When a ban on such offensive speech contains a narrow definition, it often fails for other reasons, such as viewpoint discrimination (laws that prohibit offensive speech towards one group necessarily discriminate against other groups).[45] She cites a variety of attempts to define hateful or offensive speech in other countries that similarly fail a vagueness test.[46] In sum, Article 19(3)'s "legality" condition poses a significant international legal hurdle to speech regulation that seeks to ban insulting or offensive speech.

2. The "Necessity" Condition: A Requirement to Use the Least Intrusive Means

The second prong of ICCPR Article 19(3)'s tripartite test—the "necessity" condition—is understood to require that a restriction on speech must reflect the least possible infringement on expression rights for achieving a public interest goal.[47] In other words, if a speech restriction is not the

---

43. *See id*. at 71–72 (discussing, for example, a Supreme Court case that struck down a law banning speech that undermined the "dignity" of employees at foreign embassies as inherently subjective).

44. *Id*. at 72.

45. *See id*. at 74 (illustrating that under inclusion or selectivity can be just as lethal to the legality of a speech law as vagueness). Similarly, the UN Human Rights Committee has made clear that speech restrictions must comply with other provisions in the ICCPR, including those that ban discrimination. *See* GC 34, *supra* note 35, at ¶ 26

> Laws restricting the rights enumerated in article 19, paragraph 2 . . . must not only comply with the strict requirements of article 19, paragraph 3 of the Covenant but must also themselves be compatible with the provisions, aims and objectives of the Covenant. Laws must not violate the non-discrimination provisions of the Covenant.

ICCPR Article 26 prohibits discrimination "on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." ICCPR, *supra* note 24, at art. 26.

46. *See* STROSSEN, *supra* note 42, at 75–80 (demonstrating by various examples from Canada and Europe that defining hate speech can be an intractable problem).

47. *See infra* notes 49–52 and accompanying text.

least intrusive way for achieving a governmental purpose, then it cannot be "necessary" to achieve that purpose.[48] The UN Human Rights Committee has interpreted this prong to require, *inter alia*, that the speech restriction must be "the least intrusive" means for achieving a legitimate public purpose.[49] The UN Special Rapporteur has likewise stated that governments "must demonstrate that the restriction imposes the least burden on the exercise of the right and actually protects, or is likely to protect, the legitimate State interest at issue."[50]

To make its case that it has selected the "least intrusive" speech restriction,[51] an ICCPR State Party must logically conduct a three-step inquiry:

> (1) Is there a way for the government to achieve the desired public interest goal without banning speech?
> (2) If the public purpose cannot be achieved without infringing on speech, what means are available to achieve the goal, and which one produces the "least intrusion" on speech interests?
> (3) Does the means selected actually help (or is it likely to help) achieve the public interest goal?[52]

---

48. *See infra* notes 49–52 and accompanying text.

49. *See* GC 34, *supra* note 35, at ¶ 34 (stating that speech restrictions must be "the least intrusive instrument amongst those which might achieve their protective function"). The United States government has also stated that it views the necessity condition as requiring that speech restrictions "must be the least restrictive means for protecting the governmental interest and are compatible with democratic principles." U.S. OBSERVATIONS, *supra* note 38, at 227. In addition, in order to meet the "necessity" condition, the Committee has stated that restrictions must also meet the principle of proportionality, i.e., that restrictions must "be proportionate to the interest protected." GC 34, *supra* note 35, at ¶ 34.

50. *UN SR 2018 Report*, *supra* note 34, at ¶ 7.

51. *See* GC 34, *supra* note 35, at ¶ 27 (stating that the government bears the burden of proving that each part of the ICCPR Article 19(3)'s tripartite test is met). The UN Special Rapporteur has also emphasized that governments "may not merely assert necessity but must demonstrate it, in the adoption of restrictive legislation and the restriction of specific expression." *UN SR 2018 Report*, *supra* note 34, at ¶ 7.

52. *See* Aswad, *supra* note 30, at 47 (proposing a similar three-step inquiry in the context of assessing how social media platforms that engage in content moderation could ensure they respect the "necessity" test in ICCPR Article 19(3)). The UN Special Rapporteur recently cited favorably to my

If the government can achieve a legitimate public interest goal without banning speech or through a less restrictive means, or if the selected means does not help achieve the goal, then the government has not met its burden of selecting the "least intrusive" means of attaining its goal.[53] Failure to select the "least intrusive" restriction results in the speech restriction failing the "necessity" condition.[54]

At the UN, this trilogy of questions played out in heated and divisive discussions that spanned over a decade.[55] During these discussions, diplomats contemplated whether it was appropriate to ban a spectrum of offensive speech encompassing religious intolerance/hate, blasphemy, and insults to religious feelings.[56] In 2011, at the UN Human Rights Council, the international community adopted a consensus resolution that commemorated a variety of

---

three-step inquiry in his report to the UN General Assembly, see David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 52 U.N. DOC. A/74/486 (Oct. 9, 2019) [hereinafter *UN SR 2019 Report*]

> Evelyn Aswad identifies three steps that a company should take under the necessity framework: evaluate the tools it has available to protect a legitimate objective without interfering with the speech itself; identify the tool that least intrudes on speech; and assess whether and demonstrate that the measure it selects actually achieves its goals.

These three steps—assessing if there are measures short of speech bans to achieve goals, determining the least intrusive speech infringement, and monitoring effectiveness of speech infringements—apply whether a government or company is seeking to demonstrate it acts consistently with ICCPR Article 19(3)'s necessity condition.

53. *See supra* notes 51–52 and accompanying text.

54. *See UN SR 2018 Report*, *supra* note 34, at ¶ 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.").

55. *See* Evelyn M. Aswad, *To Ban or Not to Ban Blasphemous Videos*, 44 GEO. J. INT'L L. 1313, 1323 (2013) (discussing the UN debates and discussions regarding censorship of blasphemous speech); *see also* Suzanne Nossel, ADVANCING HUMAN RIGHTS IN THE UN SYSTEM, COUNCIL ON FOREIGN REL. 15–16 (2012), https://perma.cc/P2TC-L3NC (PDF) (analyzing the debate regarding censorship of blasphemous speech in the international arena).

56. *See supra* note 55.

conclusions from these discussions.[57] While it condemns various manifestations of religious intolerance, Resolution 16/18 acknowledges that governments have a robust toolkit of options that should be deployed before resorting to speech bans.[58] For example, the resolution urges governments to, among other things, speak out against intolerance, train government employees in effective outreach measures to groups that may be subject to intolerance, promote interfaith dialogues and educational initiatives to combat intolerance, and implement robustly and fairly administered laws that ban discriminatory acts or hate crimes.[59] Moreover, the resolution noted that the freedoms of expression and religion were an important counter-weight to intolerance.[60] In addition, this resolution limited its call for speech bans to instances in which the speech rises to incitement of imminent violence—a standard that is consistent with the U.S. First Amendment.[61]

---

57. Human Rights Council Res. 16/18, U.N. Doc. A/HRC/RES/16/18 (Apr. 12, 2011) [hereinafter Council Res. 16/18].

58. *See, e.g.*, *id.* ¶ 5(h) ("Recognizing that the open, constructive and respectful debate of ideas, as well as interfaith and intercultural dialogue at the local, national and international levels, can play a positive role in combating religious hatred, incitement and violence[.]").

59. *Id.* ¶¶ 4–6.

60. *Id.* ¶¶ 4–5.

61. *See* Aswad, *supra* note 55, at 1325 (noting that even the main sponsor of the resolution recognized it was intended to be consistent with the First Amendment by citing to Ufuk Gokcen, *The Reality of Freedom of Expression in the Muslim World*, THE HILL (Oct. 19, 2012, 1:00 PM), https://perma.cc/QTW5-GFUX (last visited Jan. 7, 2020) (reflecting an acknowledgment by a representative of the Organization of Islamic Cooperation that Human Rights Council Resolution 16/18 uses "much of the United States First Amendment language" to promote "respect for and protection of the individual rights of all people") (on file with the Washington and Lee Law Review)); Jerome Socolovsky, *Islamic Nations Relinquish Demand for Defamation Laws*, VOA (Oct. 24, 2012, 10:05 PM), https://perma.cc/7QHW-EPM6 (last visited Jan. 7, 2020) (quoting a representative of the Organization of Islamic Cooperation as stating that that Resolution 16/18 protects the rights of individual believers as opposed to religion itself and the framework "is compatible with the First Amendment") (on file with the Washington and Lee Law Review). For a discussion of the First Amendment standard on incitement to imminent violence, see *infra* note 107 and accompanying text.

This resolution represents an important recognition that governments have a variety of good governance measures at their disposal to tackle intolerant, offensive, and insulting speech before claiming a need to resort to speech bans. Governments that do not implement what is known as the "16/18 consensus toolkit"[62] are hard pressed to justify their speech bans as "the least intrusive means" to achieving a legitimate purpose.[63] Prohibitions against vague speech bans and discrimination against particular groups, when coupled with a requirement to use the least restrictive means to achieve a legitimate goal, generally leave governments with limited options for banning offensive speech. All in all, ICCPR Article 19(3)'s "necessity" condition poses another significant international legal hurdle to speech restrictions on insulting or offensive speech and further supports the vision for broad freedom of expression protections expressed in *The Right to Insult in International Law*.[64]

3. The "Legitimacy" Condition: Speech Bans Must Be for Public Interest Goals

The third prong of the tripartite test—the "legitimacy" condition—requires that any speech bans be imposed for one of the legitimate, enumerated rationales set forth in ICCPR Article 19(3): "[f]or respect of the rights or reputations of others; [f]or the protection of national security or of public order (*ordre public*), or of public health or morals."[65] Both the limited number of public interest rationales in the treaty text

62. Aswad, *supra* note 55, at 1328.

63. *See id.* at 1325–26 (noting that the use of the measures in Resolution 16/18 obviates the need for "broad bans on speech").

64. For example, *The Right to Insult in International Law* advocates for speech protections "in line with the approach espoused by the U.S. Supreme Court under the First Amendment," Clooney & Webb, *supra* note 14, at 51, and Council Res. 16/18 (coupled with relevant interpretations by the UN's independent experts, see, e.g., *infra* note 90) arguably does this by calling upon states to ban offensive and hateful speech in only one instance—incitement to imminent violence, which "reflects the U.S. constitutional standard" for a permissible restriction on speech. Aswad, *supra* note 55, at 1325.

65. ICCPR, *supra* note 24, at art. 19(3).

and the UN human rights machinery's recommended interpretations serve to limit governmental discretion under this prong. The UN Human Rights Committee has warned that these grounds cannot be invoked as pretexts for limiting speech by emphasizing, for example, that "national security" cannot be improperly invoked to suppress information that does not actually harm national security.[66] The Committee has also cautioned that speech restrictions that are justified on grounds of "public morals" are only valid if judged by the "universality of human rights and the principle of non-discrimination,"[67] thereby seeking to constrain overly broad or inappropriate invocations of this rationale. Finally, the Committee has noted that the "rights . . . of others" comprises the rights in the ICCPR as well as other parts of international human rights law.[68]

The UN Human Rights Committee has criticized many national laws that seek to ban insults of rulers, religion, and royalty. For example, in its 2011 General Comment, the Committee criticized penalties for criticism or insults of public officials and royalty:

> [T]he Committee has observed that in circumstances of public debate concerning public figures in the political domain and public institutions, the value placed by the Covenant upon uninhibited expression is particularly high. Thus, the mere fact that forms of expression are considered to be insulting to a public figure is not sufficient to justify the imposition of penalties . . . . [A]ll public figures, including those exercising the highest political authority such as heads of state and government, are legitimately subject to criticism and political opposition. Accordingly, the Committee expresses concern regarding laws on such matters as, lese majesty [insulting royalty], *desacato* [criticism of rulers], disrespect for authority, disrespect for

66. GC 34, *supra* note 35, at ¶ 30.

67. *Id.* at ¶ 32. The UN Special Rapporteur reaffirmed this caution in 2018. *UN SR 2018 Report*, *supra* note 34, at ¶ 7.

68. GC 34, *supra* note 35, at ¶ 28. The UN Special Rapporteur reaffirmed this understanding in 2018. *UN SR 2018 Report*, *supra* note 34, at ¶ 7.

> flags and symbols, defamation of the head of state and the protection of the honour of public officials . . . .[69]

Similarly, the UN Special Rapporteur has taken the view that laws protecting rulers or royalty from criticism are "manifestly inconsistent with freedom of expression."[70]

The UN Human Rights Committee has also criticized bans on blasphemy or other speech that insults religions or religious sensibilities by stating the following:

> Prohibitions of displays of lack of respect for a religion or other belief system, including blasphemy laws, are incompatible with the Covenant, except in the specific circumstances envisaged in article 20, paragraph 2, of the Covenant [i.e., advocacy of incitement to harm]. . . . [I]t would be impermissible for any such laws to discriminate in favour of or against one or certain religions or belief systems, or their adherents over another, or religious believers over non-believers. Nor would it be permissible for such prohibitions to be used to prevent or punish criticism of religious leaders or commentary on religious doctrine and tenets of faith.[71]

The UN Special Rapporteur has likewise opined that religions are not entitled to reputational protections under human rights law,[72] as has the UN's top expert on freedom of religion.[73]

---

69. GC 34, *supra* note 35, at ¶ 38 (footnotes omitted).

70. *UN SR 2016 Report*, *supra* note 40, at ¶ 33.

71. GC 34, *supra* note 35, at ¶ 48.

72. U.N. SPECIAL RAPPORTEUR ON FREEDOM OF OPINION AND EXPRESSION ET AL., JOINT DECLARATION ON DEFAMATION OF RELIGIONS, AND ANTI-TERRORISM AND ANTI-EXTREMISM LEGISLATION 2 (2008), https://perma.cc/3C2L-7WWK (PDF) (noting the "defamation of religions" concept does not meet international standards for defamation, explaining that religions are not the subject of protection under these standards, and calling for an end to resolutions promoting this topic at the UN).

73. *See* Asma Jahangir (Special Rapporteur on Freedom of Religion or Belief) & Doudou Diène (Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance), *Rep. for Human Rights Council Decision 1/107 on Incitement to Racial and Religious Hatred and the Promotion of Tolerance*, ¶¶ 27, 36, 37, U.N. Doc. A/HRC/2/3 (Sept. 20, 2006) (opining that international law does not protect religions or beliefs from "criticism or ridicule").

4. Impacts on ICCPR Article 20 and CERD Article 4

The authors also express concern about the scope of ICCPR Article 20 and CERD Article 4.[74] ICCPR Article 20 requires bans on speech for "[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."[75] By its own terms, Article 20 has an intent requirement because "advocacy" necessitates an intention to promote a particular cause.[76] However, many other terms in this article have been the subject of significant debate.[77] CERD Article 4 requires states, with "due regard" to other human rights such as freedom of

74. Clooney & Webb, *supra* note 14, at 37–43.

75. ICCPR, *supra* note 24, at art. 20(2).

76. Aswad, *supra* note 55, at 1319.

77. The scope of this provision remains the subject of much debate among UN member states. For example, a 2006 report by the UN High Commissioner for Human Rights found that there was no consensus among nation states about the meaning of key terms in this article, including "incitement" and "hatred." U.N. High Comm'r for Human Rights, *Rep. on Incitement to Racial and Religious Hatred and the Promotion of Tolerance*, ¶¶ 3, 5, 81, U.N. Doc. A/HRC/2/6 (Sept. 20, 2006). The UN subsequently convened experts in four regional workshops to propose a way forward on the scope of Article 20, which culminated in the Rabat Plan of Action, but it was not endorsed by states. U.N. High Comm'r for Human Rights, *Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence,* appendix, ¶ 1, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013). Though the precise scope of Article 20 remains under debate, the application of the article is constrained by the fact that it is subject to Article 19's tripartite test. *See infra* note 80 and accompanying text. Even the negotiations about this article were contentious. Aswad, *supra* note 55, at 1320–21. Eleanor Roosevelt, the head of the U.S. delegation, fought to keep Article 20 out of the ICCPR, arguing that:

> (1) [I]t was wholly out of place in a human rights treaty to empower and require states to ban speech; (2) such a provision would be abused by repressive regimes to justify problematic crackdowns on speech; (3) the wording of the provision was ambiguous, which was very dangerous and would compound governmental abuse; and (4) the principle of democracy was better served by allowing individuals to create disputes than by suppressing speech.

Id. at 1321. Article 20 was eventually included in the ICCPR after a splintered vote. *Id.*

expression, to criminalize expressions of racial superiority and incitement to racial violence, among other forms of racial hate speech.[78]

Though it would have been preferable from a freedom of expression viewpoint if ICCPR Article 20 and CERD Article 4 had not been included in those treaties, the potential breadth of these articles is significantly limited because ICCPR Article 19(3)'s tripartite test continues to apply to restrictions imposed under these two provisions.[79] Indeed, it would be inconceivable to contemplate that either human rights treaty would empower governments to restrict speech even if the bans are vague, improperly promulgated, or unnecessary to achieve public interest purposes.

The UN treaty bodies that monitor implementation of these two treaties have made clear that any prohibitions implemented pursuant to these articles must meet ICCPR Article 19's tripartite test. For example, in General Comment 34, the UN Human Rights Committee stated, "[A] limitation that is justified on the basis of article 20 must also comply with

---

78. CERD Article 4 provides:

> States Parties . . . undertake to adopt immediate and positive 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 [which includes freedom of expression], *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 . . . .

CERD, *supra* note 25, at art. 4 (emphasis added). During the negotiations on this article, the United States expressed concerns, but eventually voted in favor of it on the basis that the "due regard" clause meant implementation of this article would be subject to protections for freedom of expression. *See* U.N. GAOR, 20th Sess., 1318th mtg. at 152, ¶ 59, U.N. Doc. A/C.3/SR.1318 (Oct. 25, 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.").

79. *See infra* notes 80–81 and accompanying text.

article 19, paragraph 3."[80] Similarly, in its General Recommendation 35, the UN CERD Committee, which is the body of independent experts elected by State Parties to monitor implementation of the CERD, took the view that any prohibitions imposed under Article 4 must comply with Article 19(3)'s conditions of legality and necessity.[81] The UN CERD Committee also stated that application of Article 4 requires "due regard" for other human rights, particularly freedom of expression, which is "the most pertinent reference principle when calibrating the legitimacy of speech restrictions."[82]

---

80. GC 34, *supra* note 35, at ¶ 50. The Committee goes on to say that in "every case in which the State restricts freedom of expression it is necessary to justify the prohibitions and their provisions in strict conformity with article 19." *Id.* at ¶ 52.

81. *See* Comm. on the Elimination of Racial Discrimination, General Recommendation No. 35, ¶¶ 12, 19, U.N. Doc. CERD/C/GC/35 (Sept. 26, 2013) [hereinafter CERD GR 35] ("The application of criminal sanctions should be governed by principles of legality, proportionality and necessity."). Presumably, the UN CERD Committee did not mention the condition of "legitimacy" because CERD Article 4 provides the legitimate public interest goal for restricting such speech. The UN CERD Committee also mentioned the principle of proportionality, which falls under the "necessity" condition. *See supra* note 49. The UN Special Rapporteur has also emphasized that any speech restrictions, including those imposed under ICCPR Article 20 or CERD Article 4, must comply with ICCPR Article 19(3)'s tripartite test. *UN SR 2019 Report*, *supra* note 52, at ¶¶ 13, 16; *UN SR 2016 Report*, *supra* note 40, at ¶¶ 17, 25.

82. CERD GR 35, *supra* note 81, at ¶ 19. The authors note that CERD Article 4's "due regard" clause has been interpreted by some to mean free expression principles must be complied with, others view it as meaning a balance must be struck between freedom of expression and the need to combat racism, and yet others believe the criminal bans on racist speech are mandatory regardless of other human rights. Clooney & Webb, *supra* note 14, at 19 n.78. General Recommendation 35 seems to espouse the first interpretation given it emphasizes compliance with ICCPR Article 19's tripartite test for laws that implement CERD Article 4 and even notes the "due regard" clause means assessing restrictions against this tripartite test. *See* CERD GR 35, *supra* note 81, at ¶ 12 (requiring consideration of the "principles of legality, proportionality and necessity"). It is unclear why *The Right to Insult in International Law* describes General Recommendation 35 as not permitting consideration of freedom of expression rights in the application of CERD Article 4. *See* Clooney & Webb, *supra* note 14, at 19 (stating that the UN CERD Committee refuses to give effect to give "due regard" to freedom of expression and mandates the criminalization of racist hate speech without regard to the legality or necessity conditions). This reading overlooks the UN CERD Committee's explicit endorsement of giving

For the reasons discussed above, the applicability of ICCPR Article 19's tripartite test to any speech bans imposed under ICCPR Article 20 or CERD Article 4 significantly curtails the potential breadth of these articles as laws implemented under such provisions must not, *inter alia*, be vague and must constitute the least intrusive means to prevent harm.[83] Though the applicability of this tripartite test to speech restrictions, including those imposed under ICCPR Article 20 and CERD Article 4, was not noted in *The Right to Insult in International Law*,[84] it significantly furthers the authors' vision for freedom of expression protections in international law.

5. Reflections on the State of UN Human Rights Law and Insulting/Offensive Speech

An analysis of ICCPR Article 19 reveals that its tripartite conditions of legality, necessity, and legitimacy are robust and

---

due regard to freedom of expression in General Recommendation 35 and its call to apply the legality and necessity principles in assessing any criminal bans imposed under Article 4. CERD GR 35, *supra* note 81, at ¶¶ 12, 19. The authors also state that a 2011 UN Special Rapporteur report takes the view that CERD Article 4 does not require consideration of expression rights. Clooney & Webb, *supra* note 14, at 19 n.80. This report, however, cited to views of the UN CERD Committee that pre-dated its 2013 General Recommendation Number 35, which reflects the UN CERD Committee's latest guidance on Article 4. *See* Frank La Rue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 31, U.N. Doc. A/66/290 (Aug. 10, 2011) (stating the Committee's view from 1993 that "the prohibition of the dissemination of all ideas based upon racial superiority or hatred is compatible with the right to freedom of opinion and expression" (quoting Comm. on the Elimination of Racial Discrimination, General Recommendation XV (42) on Article 4 of the Convention, ¶ 4, U.N. Doc. A/48 /18 (Mar. 17, 1993))).

83. *See supra* notes 24–73 and accompanying text (discussing why such conditions make it difficult for hate speech laws to pass muster under ICCPR Article 19's tripartite test).

84. *See* Clooney & Webb, *supra* note 14, at 15–19, 38–43 (discussing the criminalization of racially-motivated hate speech that appears to be required under the "plain language" of CERD Article 4, for example, but overlooking the fact that such criminalization must nevertheless satisfy the tripartite test).

constitute substantial legal hurdles for governments to surmount in order to justify any speech restrictions, including those imposed under ICCPR Article 20 and CERD Article 4.[85] Restrictions on insulting or otherwise offensive speech often fail because they are either vague or are not the least intrusive means to fulfill legitimate ends that can be achieved through good governance measures or a less onerous restriction.[86] This is fully consistent with the authors' view that "a myriad of non-legal tools" are available to combat offensive or insulting speech[87] and that criminal bans on speech may be ineffective or even counterproductive.[88] At times, restrictions on insulting speech will fail because they are imposed for an illegitimate reason, like seeking to protect rulers, religions, and royalty from insults.[89] Application of this tripartite test has put the UN human rights machinery on a path towards only permitting bans on offensive speech where there is an intent to cause harm and a likelihood of near term harm.[90]

While many states have laws that conflict with international law standards, that fact does not mean the

85. *See supra* notes 80–81 and accompanying text.

86. *See, e.g.*, *UN SR 2016 Report*, *supra* note 40, at ¶¶ 14, 25, 33 (noting examples of counter-terrorism and hate speech laws that fail ICCPR Article 19(3)'s vagueness test and highlighting that criminal penalties or inordinate civil penalties for defamation and libel are excessive, i.e., not necessary).

87. Clooney & Webb, *supra* note 14, at 41.

88. *See id*. at 43 (citing the rise of hate speech and hate crimes in Europe despite "robust" criminal laws targeting such conduct). *See also* STROSSEN, *supra* note 42, at 133–82 (explaining why hate speech laws can be ineffective and even counterproductive while non-censorial methods can be more effective in combatting intolerance).

89. *See supra* notes 69–73 and accompanying text (discussing the UN human rights machinery's criticism of laws that forbid insults to rulers, religions, or royalty).

90. *See* CERD GR 35, *supra* note 81, at ¶ 16 (noting that in assessing incitement "the intention of the speaker, and the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech in question" should be taken into account); Frank La Rue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 46, U.N. Doc. A/67/357 (Sept. 7, 2012) [hereinafter *UN SR 2012 Report*] (noting assessments of incitement should consider "real and imminent danger of violence resulting from the expression" and the speaker's intention to incite).

international treaties or the UN human rights machinery fail to adequately protect insulting or otherwise offensive speech. Rather, it means that states are either violating their ICCPR obligations or (for those not party to the ICCPR) acting inconsistently with international standards.[91] In sum, the violation of norms is a different problem from the authors' alleged insufficiency of the norms and necessitates different solutions, as discussed in Part III.

B. Why the International Law Regime Is Not Riddled with Contradictory Approaches to Insulting or Otherwise Offensive Speech

In *The Right to Insult in International Law*, the authors take the position that the international machinery charged with interpreting freedom of expression standards has issued contradictory interpretations about the permissibility of bans on insulting or otherwise offensive speech.[92] To the extent that the authors argue that the UN's *international* human rights machinery and *regional* human rights mechanisms have issued inconsistent decisions on the scope of freedom of expression, I agree.[93] Though this state of affairs is unfortunate, Part II.B.1 below explains why it does not render the UN human rights regime incoherent. Given that the UN and regional mechanisms interpret different treaties, differing approaches between the *international* and *regional* systems do not indicate fundamental flaws within the UN human rights system. To the extent that the authors argue that the various components within the UN machinery have issued inconsistent decisions about freedom of expression, Part II.B.2 below explains why the alleged contradictions in the UN's guidance

91. For example, the Human Rights Committee has expressed concern over laws that harshly penalize speech that is critical of public officials and government institutions in countries such as the Dominican Republic and Zambia. GC 34, *supra* note 35, at 9–10.

92. *See* Clooney & Webb, *supra* note 14, at 44 ("[A] close study of international legal sources reveals inconsistent guidance on the question of when international law permits insults."); *id.* at 46 ("[I]nternational law also diverges on the issue of denial laws, which protect against insults to the memories of deceased victims of crimes.").

93. *See infra* notes 94–97 and accompanying text.

are best understood as an evolution in the UN machinery's interpretations toward greater speech protections rather than as unworkable inconsistencies within the system.

1. The UN Human Rights Machinery v. Regional Human Rights Mechanisms

The authors properly observe that the UN's *international* human rights machinery, on the one hand, and *regional* human rights bodies, on the other, approach the scope of freedom of expression in different ways.[94] For example, the authors note that the European Court of Human Rights (ECtHR) has upheld the criminalization of atrocity denial under the European Convention on Human Rights (ECHR), but that the UN Human Rights Committee views such laws as incompatible with the ICCPR.[95] The authors also highlight that the Inter-American Court of Human Rights has held that, under certain circumstances, criminalization of defamation is permissible under the Inter-American Convention on Human Rights whereas the UN Human Rights Committee has taken a

94. The authors focus on the European, African, and Inter-American human rights systems in their article, *see, e.g.*, Clooney & Webb, *supra* note 14, at 24–25 n.101 (citing cases from the human rights courts in those regions), but there are other sub-global systems as well. For example, the Organization of Islamic Cooperation's human rights system is based in part on the Cairo Declaration on Human Rights in Islam that limits expression by reference to Shariah norms. Org. of Islamic Cooperation, *The Cairo Declaration on Human Rights in Islam*, Annex to Res. No. 49/19-P, art. 22 (Aug. 5, 1990), https://perma.cc/QW9C-VXGN (PDF). The Association of Southeast Asian Nations (ASEAN) also has a human rights declaration, which limits rights (including freedom of expression) in ways that are inconsistent with the UN human rights regime. *See* Press Release, U.S. Dep't of State, ASEAN Declaration on Human Rights (Nov. 20, 2012), https://perma.cc/7GYA-EYS4 (last visited Dec. 27, 2019) ("While part of the ASEAN Declaration adopted November 18 tracks the [Universal Declaration of Human Rights], we are deeply concerned that many of the ASEAN Declaration's principles and articles could weaken and erode universal human rights and fundamental freedoms as contained in the UDHR.") (on file with the Washington and Lee Law Review).

95. *See* Clooney & Webb, *supra* note 14, at 46 (comparing the views of the Human Rights Committee's GC 34, *supra* note 35, at ¶ 49, with the ECtHR's decision in Witzsch v. Germany, App. No. 7485/03, Eur. Ct. H.R., (Dec. 13, 2005), https://perma.cc/M9BR-XGSQ (PDF)).

strong stand against it when interpreting the ICCPR.[96] In addition, the authors spotlight the European Court's use of a "guillotine" provision to deny highly offensive speech any protection under the ECHR whereas the UN Human Rights Committee gives all speech the protection of ICCPR Article 19's tripartite test.[97]

Though there can be no doubt that the UN machinery and regional mechanisms have taken differing approaches concerning the scope of freedom of expression, such differences do not render the UN treaties or their associated interpretations by UN bodies inconsistent or incoherent. The international human rights law regime for freedom of expression is comprised of UN treaties (such as the ICCPR) that apply to State Parties around the world.[98] The UN's *international* human rights treaties create committees of independent experts from around the world to monitor treaty

96. *See id.* at 45–46 ("[T]he U.N. has spoken out strongly against criminal defamation . . . . [T]he Inter-American Court has held that criminal defamation is legitimate where it 'meets the requirements of necessity and proportionality' . . . ." (quoting Mémoli v. Argentina, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 265, ¶ 126 (Aug. 22, 2013))).

97. *Id.* at 22. There are many other differences between the interpretations of the UN human rights machinery and the European Court of Human Rights. For example, the UN Human Rights Committee and UN Special Rapporteur have consistently spoken out against bans on blasphemy. *See supra* notes 71–73 and accompanying text. The European Court of Human Rights, on the contrary, has consistently upheld blasphemy bans when interpreting the ECHR. *See, e.g.*, Otto-Preminger-Institut v. Austria, 295 Eur. Ct. H.R. (ser. A) (1994) (upholding prior restraint on a film that could offend the religious sensibilities of Christians); E.S. v Austria, App. 38450/12, Eur. Ct. H.R., (Oct. 25, 2018), https://perma.cc/5JDN-FN78 (last visited Dec. 28, 2019) (upholding conviction for disparaging Islamic religious doctrines) (on file with the Washington and Lee Law Review). In addition, the European Court of Human Rights has created a doctrine called the "margin of appreciation," in which it defers to the judgment of governments, particularly when there are different interpretations of treaty rights among European nations. PHILIP ALSTON & RYAN GOODMAN, INTERNATIONAL HUMAN RIGHTS 946–48 (2012). The UN Human Rights Committee, on the other hand, has made clear that it does not apply a margin of appreciation to freedom of expression issues. GC 34, *supra* note 35, at ¶ 36.

98. *See* ICCPR, *supra* note 24, at art. 48(1) (providing that any member of the United Nations and certain other states can become a party to the ICCPR).

implementation and recommend interpretations of their texts.[99] The UN Human Rights Council also periodically selects an independent expert (the Special Rapporteur) to monitor, report, and provide recommendations on global freedom of expression issues.[100] The various *regional* systems have their own treaties and monitoring and enforcement mechanisms, but they only apply within their respective areas of the world.[101] The regional human rights courts are not empowered to interpret the ICCPR; the UN system is not empowered to interpret the regional treaties. Just because regional courts and mechanisms assess their respective treaties to have a narrower scope for freedom of expression than the ICCPR[102] does not mean that the UN system—its treaties and the interpretations issued by its machinery—is somehow rendered inconsistent. It simply means that the *international* and *regional* systems provide different levels of protection for freedom of expression.

This is not to say that it is an advisable state of affairs for the international and regional systems to take inconsistent views on the scope of freedom of expression. For those of us who favor broad protections for freedom of expression, it is disheartening that regional systems provide fewer speech protections and it would be preferable if they would adopt the UN machinery's approaches. This state of affairs is also unfortunate because countries often cite to the jurisprudence

99. *See, e.g.*, *id.* art. 28(1) ("There shall be established a Human Rights Committee . . . . It shall consist of eighteen members and shall carry out the functions hereinafter provided."); CERD, *supra* note 25, at art. 8(1) ("There shall be established a Committee on the Elimination of Racial Discrimination . . . consisting of eighteen experts of high moral standing . . . .").

100. *See Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, OFF. OF THE U.N. HIGH COMMISSIONER FOR HUM. RTS., https://perma.cc/DD4N-VUXG (last visited Jan. 8, 2020) (explaining the establishment of the Special Rapporteur's mandate in 1993 and the periodic extensions thereof) (on file with the Washington and Lee Law Review).

101. *See, e.g.*, Convention for the Protection of Human Rights and Fundamental Freedoms, *opened for signature* Nov. 4, 1950, E.T.S. NO. 005, art. 59 (establishing that only members of the Council of Europe can become state parties to the Convention).

102. *See supra* notes 95–97 and accompanying text.

of their regional machinery to justify their actions when they are accused of violating UN human rights treaties. For example, when the UN Special Rapporteur wrote to Germany to express a host of serious concerns with its Network Enforcement Act,[103] Germany defended its law based on European Court of Human Rights jurisprudence.[104] Though invocation of a *regional* treaty cannot justify a country's violation of its *international* treaty obligations,[105] such attempts may prove dangerous over time as they could eventually erode UN protections for speech when ICCPR State Parties invoke regional norms and worldviews to advocate for lower speech protection standards at the international level.

2. Unworkable Inconsistencies Within the UN's Human Rights Machinery or an Evolution of Interpretative Guidance?

The authors also take the position that the UN human rights machinery's guidance on the scope of protection for freedom of expression is internally inconsistent. This Part posits that such differences would be better characterized as an evolution in the thinking of these bodies, with more recent interpretations reflecting a progression towards broader protections for speech. Such an evolution in legal interpretation happens as well within domestic court systems. For example, U.S. Supreme Court jurisprudence once upheld bans on speech that had a "bad tendency" to cause harm.[106]

---

103. Netzwerkdurchsetzungsgesetz [NetzDG] [Network Enforcement Act], Sept. 1, 2017, BUNDESGESETZBLATT, Teil I [BGBL I] at 3352 (Ger.), https://perma.cc/V46W-J3A4 (PDF).

104. *See* Letter from Federal Government of Germany, Answers to the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression in Regard to the Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act) 3 (June 1, 2017), https://perma.cc/95YG-MKAK (PDF) ("Likewise, the European Court of Human Rights in its case-law has made abundantly clear that hate speech is intolerable in a democratic society.").

105. As noted by the UN Special Rapporteur, "[r]egional human rights norms cannot, in any event, be invoked to justify departure from international human rights protections." *UN SR 2019 Report*, *supra* note 52, at ¶ 26.

106. *See* ANTHONY LEWIS, FREEDOM FOR THE THOUGHT THAT WE HATE: A BIOGRAPHY OF THE FIRST AMENDMENT 23–38, 157–167 (2007) (noting that this

Over time, however, it shifted towards a more rigorous standard that requires the speaker's intent to cause lawless action that is both likely and imminent.[107] This Part analyzes the authors' main examples of the UN machinery's alleged inconsistencies and proposes that these decisions should instead be viewed as an evolution in the thinking of these bodies, which appear to be on a trajectory towards greater protections for speech.

As evidence of an inconsistency within the UN human rights machinery, the authors note the UN CERD Committee's differing views as to what constitutes "incitement."[108] In particular, the authors highlight that the UN CERD Committee's General Recommendation 35 states that incitement should include the "imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech . . . ."[109] The authors then contrast this standard with individual complaints decisions issued by the UN CERD Committee and UN Human Rights Committee.[110] However, the cited individual complaints decisions were all

---

view prevailed in early twentieth-century cases such as Patterson v. Colorado, 205 U.S. 454 (1907)).

107. *See id.* at 157–67 (discussing the ruling in *Brandenburg v. Ohio*, 395 U.S. 444 (1969)).

108. *See* Clooney & Webb, *supra* note 14, at 36 (comparing the CERD Committee's General Recommendation 35 with its decisions in individual complaints as well as the Human Rights Committee's General Comment 34).

109. *Id.* (quoting CERD GR 35, *supra* note 81, at ¶ 16).

110. *See id.* ("[I]n many cases both the Human Rights Committee and the CERD Committee have sanctioned criminal convictions for speech without explicitly taking into account whether a prohibited impact was foreseen, likely or imminent."). The UN's Human Rights Committee and CERD Committee can hear individual complaints against States Parties for treaty violations if those governments have consented to subject themselves to such complaints. *See Human Rights Treaty Bodies—Individual Communications*, OFF. OF THE U.N. HIGH COMMISSIONER FOR HUM. RTS., https://perma.cc/4YZY-UNWQ (last visited Jan. 8, 2020) ("Anyone can lodge a complaint with a Committee against a State that satisfies these two conditions (being a party to the treaty and having accepted the Committee's competence to examine individual complaints), claiming that his or her rights under the relevant treaty have been violated.") (on file with the Washington and Lee Law Review).

decided before CERD General Recommendation 35 was issued in 2013.[111]

This General Recommendation should be viewed as an evolution in the Committee's thinking on bans on racist hate speech rather than an inconsistency in its jurisprudence. Before issuing formal recommended interpretations of treaties, UN treaty committee members engage in a consultation process that canvases the views of State Parties and civil society organizations.[112] After such consultations, committees hold intensive deliberations that can take years before issuing the committee's ultimate views.[113] Such a period of intensive deliberation and reflection can easily mark a turning point and evolution in the UN 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.[114]

The authors also express concern about potential contradictions between the UN Human Rights Committee and the UN CERD Committee with respect to the criminalization

111. *See* Clooney & Webb, *supra* note 14, at 36 n.165 (citing UN Human Rights Committee decisions from 1996, 2000, and 2009 as well as a UN CERD Committee decision from 2005).

112. *See, e.g.*, Michael O'Flaherty, *Freedom of Expression: Article 19 of the International Covenant on Civil and Political Rights and the Human Rights Committee's General Comment 34*, 12 HUM RTS. L. REV. 627, 650 (2012) (discussing submissions by States Parties and other stakeholders suggesting changes to a draft of General Comment 34).

113. *See id.* at 645 (describing the process by which the UN Human Rights Committee developed General Comment 34).

114. On the topic of inconsistencies on "incitement" by the UN machinery, the authors also state that a draft of the UN Human Rights Committee's General Comment 34 contained language that stated Article 20 required advocacy that was "likely to trigger imminent acts," but that such guidance was not included in the final version of the General Comment. Clooney & Webb, *supra* note 14, at 36. The lead drafter on the UN Human Rights Committee for this 2011 General Comment has explained that "he had not been mandated to develop general comment guidance on [Article 20]" and was only able to reach consensus on text that explained the relationship between Articles 19 and 20. O'Flaherty, *supra* note 112, at 647–49. Thus, removal of the draft language interpreting Article 20 may have reflected a need to stay within the mandate to develop guidance on Article 19 rather than a definitive, substantive stance by the UN Human Rights Committee on the drafter's proposed language interpreting Article 20.

of defamation.[115] The authors note that in 2011 the UN Human Rights Committee had "spoken out strongly against criminal defamation" in both an individual complaints decision and its General Comment 34,[116] but cite to a 2003 UN CERD Committee individual complaints decision, in which it upheld a criminal defamation conviction in Denmark, as evidence of being out of step with its sister committee.[117] This decision was issued well before the UN CERD Committee's 2013 General Recommendation 35, which sets forth how the Committee will analyze the criminalization of such speech applying the conditions of legality and necessity in future cases.[118] Might

115. Clooney & Webb, *supra* note 14, at 44–45.

116. *Id.* at 45 n.203, 204. The article also notes that the UN Special Rapporteur "consistently advocates the decriminalization of defamation. . . . [but] has, on occasion, implicitly regarded criminal defamation laws as acceptable." *Id.* at 45 n.205. The article cites as an example a 2016 letter of the UN Special Rapporteur to the Netherlands about his concern and criticism of the country's *lese-majeste* [insulting royalty] laws. *Id.* It is unclear why the article views this letter as an implicit approval of criminal defamation laws. In this letter, the Special Rapporteur condemns such laws as not proportional (i.e., they fail ICCPR Article 19's tripartite test) and cites to the UN Human Rights Committee's criticisms of criminal defamation in General Comment 34. Letter from David Kaye, U.N. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, to Roderick van Schreven, Permanent Representative of the Netherlands to the United Nations, 3–4 (Oct. 14, 2016), https://perma.cc/4ZQZ-VKNA (PDF). This footnote also cites to an opinion by Europe's Commission for Democracy through Law (also known as the Venice Commission). Clooney & Webb, *supra* note 14, at 45 n.205. However, the Venice Commission is not part of the UN's machinery (instead being part of Europe's regional human rights system) and thus not relevant to the argument that the guidance issued by the UN machinery is internally inconsistent. Moreover, this Venice Commission decision re-states General Comment 34's criticism of criminal defamation laws. *See* Venice Comm'n, *Opinion on Articles 216, 299, 301 and 314 of the Penal Code of Turkey*, Opinion No. 831/2015, ¶ 56 (Mar. 15, 2016), https://perma.cc/3GS5-DBYS (PDF) ("Under Article 19 ICCPR, the United Nations Human Rights Committee has also urged that 'States parties should consider the decriminalization of defamation . . . and *imprisonment is never an appropriate penalty*." (quoting GC 34, *supra* note 35, at ¶ 47)).

117. Clooney & Webb, *supra* note 14, at 45 n.206.

118. *See supra* notes 81–82 and accompanying text. Moreover, the 2003 CERD individual complaints decision also pre-dates the UN Human Rights Committee's General Comment 34, which was issued in 2011. GC 34, *supra* note 35. Often the UN CERD Committee will consider the views of its sister

not this chronological sequencing of events be better viewed as an evolution in the guidance of the UN bodies towards greater protections for speech rather than an unworkable inconsistency?

The authors also argue that the UN Human Rights Committee and the UN CERD Committee have at times suggested that State Parties must satisfy an intent requirement before banning incitement to discrimination, hostility, or violence and at other times seem to have approved of such bans without insisting on such a showing.[119] For example, the authors highlight the UN CERD Committee's General Recommendation 35, which states that any convictions for incitement should satisfy an intent standard.[120] The article then cites to a variety of UN Human Rights Committee and UN CERD Committee individual complaints decisions that "have allowed convictions for hate speech to stand even where there was no analysis of intent . . . ."[121] All of the individual complaints decisions cited to for this proposition pre-date the UN Human Rights Committee's General Comment 34 (from 2011) and the UN CERD Committee's

committee, the UN Human Rights Committee, in its work. *See, e.g.,* CERD GR 35, *supra* note 81, at ¶¶ 4, 6 (highlighting the UN Human Rights Committee General Comment 34 in noting its use of sister committees' interpretations). The article also cites to UN CERD Committee individual complaints decisions from April 15, 2005 and April 4, 2013. Clooney & Webb, *supra* note 14, at 45 n.208. Both decisions pre-dated the UN CERD Committee's General Recommendation 35, which was issued in September 2013.

119. *See* Clooney & Webb, *supra* note 14, at 33 ("Nevertheless, both the CERD Committee and the Human Rights Committee have allowed convictions for hate speech to stand even where there was no analysis of intent, including in recent years." (footnotes omitted)).

120. *See id.*

> [I]n its 2013 General Recommendation on racial hate speech the CERD Committee published new guidance that State parties should recognize as "important elements" of any offence of incitement "the intention of the speaker and the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech in question."

121. *Id.*

General Recommendation 35 (from 2013).[122] Again, when one examines the chronology of the cited decisions and the overall trajectory towards greater protections for freedom of expression, it seems more appropriate to view this sequencing as an evolution in the UN machinery's thinking rather than as an unworkable or problematic inconsistency.

3. Reflections on Alleged Inconsistencies in UN Freedom of Expression Interpretations

An analysis of the alleged inconsistencies in the international law regime results in two main conclusions. First, although the UN's *international* human rights regime and the various *regional* human rights regimes approach freedom of expression differently, that fact does not somehow render the UN regime incoherent or inconsistent.[123] It simply means that the UN system provides more protections for speech than the regional systems.[124] This state of affairs risks eroding UN standards over time when governments improperly invoke their regional standards for freedom of expression as an excuse for not implementing their international human rights obligations or assert that regional interpretations can define the scope of international human rights obligations.[125] Unfortunately, collapsing the UN and regional systems to conclude that the "international" system is incoherent only strengthens the potential for such erosion because it incentivizes countries that oppose UN standards to counter them by creating and invoking "competing" regional norms.

Second, based on the specific arguments made in *The Right to Insult in International Law*, it appears that the alleged inconsistencies within the UN machinery are more likely evolutions in the thinking of the UN mechanisms rather

122. *See id.* at 33 nn.148–49 (citing individual complaint decisions from 1996, 2003, 2005, and 2009).

123. *See supra* Part II.B.1.

124. *See supra* notes 94–97 and accompanying text.

125. *See supra* notes 104–105 and accompanying text.

than problems that render their interpretations incoherent.[126] Though the article usually contrasted earlier decisions of the UN's oversight machinery with its most recent decisions and commentary to find inconsistencies, the sequencing of the cited decisions and interpretations reflect an evolution towards greater speech protections. The United States domestic legal system underwent a similar evolution, with the Supreme Court gradually expanding protections for speech over time. The UN's trajectory towards greater recognition of speech protections should be reinforced rather than dismissed as unworkable inconsistencies.

III. Assessment of Proposals to Promote Free Expression at the International Level

The authors make four recommendations for promoting broad protections for insulting or offensive speech at the international level, including by updating the existing international regime or seeking particular interpretations of existing provisions.[127] The authors conclude by calling on

126. The point of this analysis is not to argue that the UN machinery's interpretations are perfect or always consistent, but rather to address the particular claims about systemic inconsistences that are made in *The Right to Insult in International Law*, which (as discussed in Part I) has resulted in the article being cited for the proposition that the UN human rights regime for freedom of expression may be problematic in the context of content moderation by global platforms. *See supra* note 13 and accompanying text.

127. *See* Clooney & Webb, *supra* note 14, at 53–54 (proposing that CERD Article 4 be deleted or "excluded through reservations," that "states should enter reservations to ICCPR Article 20," that outdated concepts not be utilized "to justify the criminalization of insults," and that ECHR Article 17 and ICCPR Article 5 should be made inapplicable to "freedom of expression cases"). The authors also set forth recommendations for action at the domestic level, including: the recognition in local laws of a "right to insult" unless the expression intends to create violence or other crimes and is likely to do so; the abolition of criminal laws on insulting rulers, religion, and royalty as well as bans on denials of historic atrocities; and an end to misusing counter-terrorism or public order laws to ban insulting speech. *Id.* at 52–53. I generally agree with the substance of these recommendations. However, as previously discussed in Part II, these recommendations already align with existing UN treaty texts and reasonable interpretations of those treaties by the UN's human rights machinery. *See supra* notes 66–84 and accompanying text. The reason countries have problematic laws that ban insulting rulers, religion, or royalty is not because international law

private platforms to be guided by international human rights law, as amended by their recommendations, in curating online content.[128] This Part evaluates each proposal, determines that the international regime is much closer to the authors' vision than they had assessed, and proposes more impactful ways of achieving the authors' important goal of promoting broad speech protections at the international level.

A. CERD Article 4

The authors' first recommendation is that "CERD Article 4 should be deleted by the agreement of States Parties or excluded through reservations."[129] Although the CERD does provide an amendment procedure,[130] given the alarming global

---

standards are lacking but because the countries lack the political will to change their laws and nations have not afforded effective enforcement mechanisms for international human rights treaties. *See* Alberto Cerda Silva, *Protecting Free Speech in the Digital Age: Q&A with UN Special Rapporteur for Freedom of Expression*, FORD FOUNDATION (Oct. 11, 2016), https://perma.cc/9LCJ-QXE9 (last visited Jan. 7, 2020)

> Around the world, governments are stretching the meaning of what's permitted [under ICCPR Article 19]: They define national security in broad, vague ways that make it difficult for individuals to know what speech or opinions are allowed and what may be subject to penalty, they apply restrictions that go well beyond what is necessary to address specific threats, and they fail to justify their restrictions.

(on file with the Washington and Lee Law Review). It is therefore important, as I argue later in this Article, for human rights advocacy to invoke and promote ICCPR Article 19's tripartite test and the most recent guidance from the UN's human rights machinery in order to protect freedom of expression at the international level. *See infra* Part III.E.

128. *See* Clooney & Webb, *supra* note 14, at 54 (advocating for technology companies to use international human rights law as well as the authors' suggestions as guides for determining whether online content should be removed).

129. Clooney & Webb, *supra* note 14, at 53.

130. The CERD requires that a State Party submit any request for revision of the treaty to the UN Secretary General and that the UN General Assembly decide how to proceed. *See* CERD, *supra* note 25, at art. 23

> (1) A request for the revision of this Convention may be made at any time by any State Party by means of a notification in writing addressed to the Secretary-General of the United Nations. (2) The

trend toward greater governmental repression of speech,[131] it is not likely that opening up a negotiation among states on CERD Article 4 would result in its deletion rather than further restrictions on speech, especially because governments tend to fight for text that justifies their existing domestic practices in multilateral human rights fora.[132] For example, for many years, the Organization of Islamic Cooperation—which represents fifty-seven nations—led an effort at the UN to develop a protocol to the CERD that would have further restricted speech by banning the "defamation of religions," an umbrella concept that includes blasphemy, insults to religions, and other speech that offends religious sensibilities.[133] A

> General Assembly of the United Nations shall decide upon the steps, if any, to be taken in respect of such a request.

In a prior attempt to amend the CERD in the early 1990s, the General Assembly approved the proposed amendment subject to a proviso that it would not go into force until two-thirds of the State Parties had endorsed it, which has not yet occurred. *See* Amendment to Article 8 of the International Convention on the Elimination of All Forms of Racial Discrimination 1, United Nations Treaty Collection (Jan. 15, 1992), https://perma.cc/7LL4-939H (PDF) (documenting the status of the proposed amendment).

131. As the authors noted, "the tide around the world is turning against free speech . . . ." Clooney & Webb, *supra* note 14, at 55. *See also Democracy in Retreat: Freedom in the World 2019*, 5 FREEDOM HOUSE, https://perma.cc/PRA3-JVRE (last visited Dec. 29, 2019) (finding that "data show freedom of expression declining each year over the past [thirteen] years, with sharper drops since 2012") (on file with the Washington and Lee Law Review). For an overview of the growing European restrictions on speech, see David Kaye, *How Europe's New Internet Laws Threaten Freedom of Expression: Recent Regulations Risk Censoring Legitimate Content*, FOREIGN AFF. (Dec. 18, 2017), https://perma.cc/SU9Y-8TE5 (last visited Dec. 29, 2019) (explaining European countries' attempts to regulate online speech and the impact that has on private internet companies and individuals) (on file with the Washington and Lee Law Review). *See also Dangerously Disproportionate: The Ever-Expanding National Security State in Europe*, AMNESTY INT'L (Jan. 17, 2017), https://perma.cc/GAV5-6QPQ (last visited Mar. 18, 2020) (reporting on European governments' responses to national security threats and the adverse effect of those responses on a variety of rights including freedom of expression) (on file with the Washington and Lee Law Review).

132. *See infra* notes 133–134 and accompanying text.

133. *See* Robert C. Blitt, *Defamation of Religion: Rumors of Its Death Are Greatly Exaggerated*, 62 CASE WESTERN RES. L. REV. 347, 365–66 (2011) (discussing how a committee was formed at the UN to develop a protocol to the CERD that would have criminalized blasphemy). Freedom House and members of civil society expressed alarm at the attempt by states at the UN

sustained diplomatic effort by many countries, including the United States, and civil society avoided that outcome, aligned the discussions with international human rights law, and shifted the focus to measures that would combat religious intolerance and negative stereotyping without banning speech.[134] Amending existing human rights treaties during a time of rising governmental restrictions on speech would not bode well for achieving deletion of CERD Article 4 or the larger goal of promoting the broadest possible speech protections. Instead, at any negotiation to amend the CERD, governments would more likely seek to commemorate text that justifies their problematic laws, including those that the authors highlight in their survey of worldwide bans on insults against rulers, religion, and royalty.[135]

The authors' recommendation provides in the alternative that nations should take a reservation to CERD Article 4.[136]

---

to amend the CERD to criminalize the "defamation of religions." *See* FREEDOM HOUSE, FACT SHEET: DEFAMATION OF RELIGIONS 1–2 (2009), https://perma.cc/3GDT-HQY4 (PDF) (explaining why "the concept of 'defamation of religions'" is problematic); L. Bennett Graham, *No to an International Blasphemy Law*, THE GUARDIAN (Mar. 25, 2010, 11:30 EDT), https://perma.cc/AQ37-4RRE (last visited Dec. 29, 2019) (arguing against the adoption of an international treaty banning blasphemy on the grounds that the human rights regime is intended to protect human beings and not ideas) (on file with the Washington and Lee Law Review).

134. *See* Nossel, *supra* note 55, at 15–16 (detailing the role of the United States in dissuading the passage of a defamation of religion protocol at the UN). The risk of re-igniting the process to develop a protocol to the CERD to criminalize the "defamation of religions" remains and freedom of expression advocates have been vigilant in fighting against such attempts to re-open speech restrictions in the CERD. *See UN HRC: Initiative to Criminalise "Acts of a Racist and Xenophobic Nature" Must Adhere to Freedom of Expression Standards,* ARTICLE 19 (Mar. 21, 2017), https://perma.cc/CD2X-TJ26 (last visited Dec. 29, 2019) (urging states at the Human Rights Council to "to guard against any new binding instrument that may be used as a vehicle to undermine protections for freedom of expression and equality") (on file with the Washington and Lee Law Review).

135. *See* Clooney & Webb, *supra* note 14, at 3–13 (reviewing and categorizing various countries' laws on "insulting rulers," "insulting religion," and "insulting royalty").

136. *See* Clooney & Webb, *supra* note 14, at 53 ("CERD Article 4 should be deleted by the agreement of States Parties or excluded through reservations"). The Vienna Convention on the Law of Treaties defines a reservation as "a unilateral statement . . . made by a State, when signing,

But it is questionable how helpful this strategy would be in promoting broad international legal protections for freedom of expression. Although the CERD has a procedure for lodging reservations, those reservations must be taken at the time when a state becomes a party to a treaty.[137] Because the CERD already has 182 State Parties,[138] none of those existing State Parties are able to take a reservation. As a result, only a dozen or so states could even take the recommended reservation to Article 4 upon joining the treaty.[139]

While new State Parties taking a reservation to CERD Article 4 would have the benefit of meaning that they refuse to take on a treaty obligation to ban speech, it is unclear how helpful a series of such reservations would be with respect to the state of international law on mandatory racist hate speech bans. Reservations that decry or commemorate the broadness of CERD Article 4 would have the (unintended) consequence of bolstering existing State Parties claims that they have expansive rights to ban speech under this provision.[140] It may

---

ratifying, accepting, approving or acceding to a treaty, whereby it purports to exclude or to modify the legal effect of certain provisions of the treaty in their application to that State." Vienna Convention on the Law of Treaties, art. 2(1)(d), May 23, 1969, 1155 U.N.T.S. 331 [hereinafter VCLT].

137. *See* CERD, *supra* note 25, at art. 20(1) ("The Secretary-General of the United Nations shall receive and circulate to all States which are or may become Parties to this Convention reservations made by States *at the time of ratification or accession*.") (emphasis added). This timing requirement is consistent with the definition of a "reservation" in the Vienna Convention on the Law of Treaties. *See* VCLT, *supra* note 136, at art. 2(1)(d) (stating that a reservation is made at the time when a State is "signing, ratifying, accepting, approving, or acceding to a treaty").

138. *See International Convention on the Elimination of All Forms of Racial Discrimination*, UNITED NATIONS TREATY COLLECTION, https://perma.cc/Y4CM-L67X (last visited Dec. 29, 2019) (listing the countries that are state parties to the CERD) (on file with the Washington and Lee Law Review).

139. The UN member states that have not yet become State Parties to the CERD include Bhutan, Brunei Darussalam, North Korea, Kiribati, Malaysia, Micronesia, Myanmar, Nauru, Palau, Samoa, South Sudan, Tuvalu, and Vanuatu. *See International Convention on the Elimination of All Forms of Racial Discrimination*, *supra* note 138 (indexing the countries that have become state parties to the CERD).

140. *See* Clooney & Webb, *supra* note 14, at 19 n. 78 (noting that one of the three main ways of interpreting CERD Article 4 is "that the protection of

therefore be preferable for non-parties to the CERD (if they have the political will to do so) to issue an "understanding"[141] that recognizes that Article 4 is subject to ICCPR Article 19's tripartite test, rather than a reservation, when joining the CERD. This would help solidify and give impetus to a narrower view of the scope of CERD Article 4, and thus better achieve the authors' goal because understandings "can be important when interpreting the meaning of a treaty since, if other states do not object or offer alternative ones, the understandings might be viewed as evidence of the parties' interpretation of the treaty."[142]

The authors' recommendation on CERD Article 4 concludes by asserting that its breadth "cannot be cured by interpretation."[143] As analyzed in Part II, rigorous application of ICCPR Article 19's tripartite test to CERD Article 4 significantly narrows the potential breadth of this article because governments must prove, *inter alia*, that any speech bans are not vague and constitute the least intrusive means to fight racism, which are both significant hurdles that narrow appropriate invocations of this article to justify speech bans.[144] The UN CERD Committee's General Recommendation 35 similarly provides that imposition of hate speech bans under CERD Article 4 must meet the conditions of legality, necessity,

---

human rights may not be invoked to avoid enacting legislation to give effect to" this ban on racist hate speech by citing to Karl Josef Partsch, *Racial Speech and Human Rights: Article 4 of the Convention on the Elimination of All Forms of Racial Discrimination*, *in* STRIKING A BALANCE: HATE SPEECH, FREEDOM OF EXPRESSION AND NON-DISCRIMINATION 24–25 (Sandra Coliver et al. eds., 1992)).

141. At the time when a state joins a treaty it can issue an "understanding," which "is an interpretative statement that clarifies or elaborates upon a treaty provision without altering it." SEAN MURPHY, PRINCIPLES OF INTERNATIONAL LAW 94 (West Acad. 2d ed. 2018). Understandings are therefore different from "reservations," which allow states to exclude or modify treaty obligations. *See* VCLT, *supra* note 136, at art. 2(1)(d), (defining a "reservation" as a unilateral statement made by a country that "purports to exclude or to modify the legal effect of certain provisions of the treaty in their application to that State").

142. MURPHY, *supra* note 141, at 94.

143. Clooney & Webb, *supra* note 14, at 54.

144. *See supra* notes 24–71 and accompanying text.

and proportionality.[145] In addition, General Recommendation 35 views incitement as limited to cases in which there is sufficient intent and an imminent risk or likelihood of harm, which further gives momentum to the authors' goal of limiting the potential breadth of this article.[146] Rather than taking the position that this article cannot be narrowed by interpretation or promoting reservations that commemorate an undesirable breadth for this article, why not steal victory (broad speech protections) from the jaws of potential defeat (CERD Article 4) by emphasizing and promoting several key interpretations of the UN CERD Committee in human rights advocacy?

B. ICCPR Article 20

The authors' second recommendation is for states to "enter reservations to ICCPR Article 20 to prohibit speech only where it intentionally incites violence or a criminal offence that is likely to follow imminently (or is otherwise concretely identified) as the result of the speech."[147] Because there are already 173 State Parties to the ICCPR[148] and reservations cannot be lodged after a state joins a treaty, this recommendation would only apply to those few states that are not already parties to the treaty.[149]

This recommendation triggers some of the same concerns noted above with respect to the proposal to promote reservations to CERD Article 4. First, it is unclear that ICCPR non-parties with poor records on human rights (including with respect to freedom of expression) will join the ICCPR with the

145. *See supra* note 81 and accompanying text.

146. *See supra* note 109 and accompanying text.

147. Clooney & Webb, *supra* note 14, at 54.

148. *See International Covenant on Civil and Political Rights,* UNITED NATIONS TREATY COLLECTION, https://perma.cc/CE2N-7J2T (last visited Dec. 29, 2019) (listing the countries that are signatories and state parties to the ICCPR) (on file with the Washington and Lee Law Review).

149. China, Cuba, Comoros, Nauru, Palau, and Saint Lucia have signed, but not ratified the ICCPR. *See id.* (indexing the countries that are party to the ICCPR). Additional UN member states that are not State Parties to the ICCPR include Bhutan, Brunei, Kiribati, Malaysia, Micronesia, Myanmar, Oman, Saint Kitts and Nevis, Saudi Arabia, Singapore, Solomon Islands, South Sudan, Tuvalu, and the United Arab Emirates. *Id.*

proposed reservation. Second, if future parties to the ICCPR have the political will to speak out on this issue, it would be more strategic to encourage them to join with "understandings" that commemorate the substance of the authors' suggested interpretation for the reasons articulated in the prior discussion on CERD Article 4.[150] While a series of reservations could unfortunately help to solidify an interpretation of Article 20 that permits broad speech bans, the issuance of "understandings" that commemorate the UN Human Rights Committee and UN Special Rapporteur's interpretations of Articles 19 and 20 would be more productive because they would indicate support for a more limited scope for this article.

An understanding, for example, could note that any restriction imposed under Article 20 must pass Article 19's tripartite test, as the UN Human Rights Committee recommends.[151] Such an understanding would build momentum to limit the potential scope of Article 20.[152] As analyzed previously, it will be difficult for bans on offensive speech to pass the legality condition as many restrictions will be vague.[153] It will also be very difficult for a government to prove it does not have any good governance means except a broad ban on speech to achieve a legitimate objective when the offensive speech is not linked to violence or other harms in the near term.[154] The understanding could also highlight that ICCPR Article 20 contains an intent element and requires a "real and imminent" threat of harm, as the UN Special Rapporteur noted.[155] Again, understandings that commemorate the substance of the UN human rights machinery's recommended interpretations[156] would help

150. *See supra* notes 136–142 and accompanying text.

151. *See supra* note 80 and accompanying text.

152. *See supra* notes 80–82 and accompanying text.

153. *See supra* notes 37–46 and accompanying text.

154. *See supra* notes 47–64 and accompanying text.

155. *See UN SR 2012 Report*, *supra* note 90, at ¶ 46 (explaining that "the "intent of the speaker to incite discrimination, hostility or violence" and the "real and imminent danger of violence resulting from the expression" are key factors in assessing whether speech incites harms).

156. *See supra* notes 31–83 and accompanying text.

achieve the authors' goal of promoting broad speech protections. Why not steal victory (broad speech protections) from the jaws of potential defeat (ICCPR Article 20) by emphasizing the UN machinery's recommended interpretations and promoting widespread acceptance and implementation of these existing interpretations?

C. "Vague" and "Anachronistic" Reasons for Governmental Speech Restrictions

The authors express concerns that terms such as "public morals" in ICCPR Article 19(3) and "propaganda for war" in ICCPR Article 20 are vague and "should not be relied on to justify the criminalization of insults."[157] They recommend such terms "should be defined more precisely" or "states should enter reservations or declarations in respect of these terms."[158] If the "defining" of these terms implicates re-negotiating the ICCPR text to include definitions, such a course of action is likely to result in a roll-back for freedom of expression rather than a narrowing of these terms for the reasons discussed above.[159] The lodging of reservations by future ICCPR State Parties would also likely backfire as reservations lamenting the breadth of these terms would give momentum to existing State Parties' claims about the broad scope of governmental prerogatives under Article 20.[160]

Issuance of understandings or, as the authors suggest, "declarations"[161] would be preferable to a reservation, as previously discussed.[162] Such understandings or declarations

157. Clooney & Webb, *supra* note 14, at 54. ICCPR Article 19(3) lists "public morals" as one of the reasons that governments can invoke to restrict speech. *See* ICCPR, *supra* note 24, at art. 19(3) ("The exercise of the rights . . . may therefore be subject to certain restrictions . . . [f]or the protection of . . . public . . . morals."); *see also id.* art. 20(1) ("Any propaganda for war shall be prohibited by law.").

158. Clooney & Webb, *supra* note 14, at 54.

159. *See supra* notes 131–134 and accompanying text.

160. *See supra* notes 140–142 and accompanying text.

161. MURPHY, *supra* note 141, at 94 ("A declaration is a statement expressing the state's position or opinion on matters relating [to] the treaty, such as whether to accept an optional form of binding dispute resolution.").

162. *See supra* notes 136–142 and accompanying text.

could cite to the substance of existing UN human rights machinery definitions of these terms that have sought to limit their scope. For example, the UN Human Rights Committee sought to narrow the potential discretion of governments when they seek to justify speech bans based on "public morals" when it took the position that "'the purpose of protecting public morals must be based on principles not deriving exclusively from a single tradition. Any such limitations must be understood in the light of universality of human rights and the principle of non-discrimination."[163] Referring to such interpretations in future understandings or declarations could give additional momentum to a restrictive scope to the grounds set forth in the "legitimacy" prong of ICCPR Article 19.[164]

D. The Rights of Others

The authors' last recommendation at the international level is that ECHR Article 17[165] and ICCPR Article 5[166] "should not apply to freedom of expression cases, which should always be subject to a balancing test under Articles 10 and 19 of these conventions, respectively."[167] ECHR Article 17 and ICCPR

---

163. GC 34, *supra* note 35, at ¶ 32.

164. *See supra* notes 140–142 and accompanying text.

165. *See* Convention for the Protection of Human Rights, *supra* note 101, at art. 17

> Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention.

166. *See* ICCPR, *supra* note 24, at art. 5

> 1. Nothing in the present Covenant may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms recognized herein or at their limitation to a greater extent than is provided for in the present Covenant. 2. There shall be no restriction upon or derogation from any of the fundamental human rights recognized or existing in any State Party to the present Covenant pursuant to law, conventions, regulations or custom on the pretext that the present Covenant does not recognize such rights or that it recognizes them to a lesser extent.

167. Clooney & Webb, *supra* note 14, at 54.

Article 5 provide that the respective treaties do not authorize activities or acts aimed at destroying the rights in the treaty.[168] As the authors note, the European Court of Human Rights has interpreted ECHR Article 17 to mean that certain types of highly offensive speech are outside of the scope of the ECHR and thus do not enjoy protection under Article 10.[169] The authors also highlight that, although ICCPR Article 5 is analogous to ECHR Article 17, the UN Human Rights Committee applies Article 19's tripartite test to all freedom of expression cases.[170]

While I fully agree with the authors that ECHR Article 17 and ICCPR Article 5 should not be applied in a fashion that prevents application of their respective convention's speech protections, this problem exists solely within Europe's regional human rights system and not in the international system. Though it is of course important to remain vigilant that such a concept not migrate from the *regional* ECHR jurisprudence into the *international* human rights system, it is unclear that any action is needed with respect to the *international* system on this topic. It is also important for human rights advocates and scholars to encourage the European Court of Human Rights to cease its guillotine-like application of Article 17 and instead assess the validity of all restrictions on speech.[171] Again, acknowledging and promoting the existing interpretations and practices at the UN level on this topic and advocating changes to the regional system is preferable to implying that European interpretations of European regional

168. *See supra* notes 165–166 and accompanying text.

169. *See* Clooney & Webb, *supra* note 14, at 22

> Article 17 has been interpreted [by the European Court of Human Rights] to mean that speech which is so odious that it could not possibly be protected under Article 10 of the Convention can be dealt with under Article 17, which allows a case to be struck out without examination of the merits.

170. *See id.* at 22 n.92 ("Article 5 of the ICCPR is analogous to Article 17 of the ECHR, but the Human Rights Committee has not used it to strike out cases without considering the merits under Article 19 or 20 of the ICCPR.").

171. *See id.* at 22 (explaining how ECHR Article 17 functions as a guillotine-like "provision because it does not involve any balancing of the right to free expression against the other values protected in Article 10").

human rights instruments create problems within the UN's international human rights system.

E. Reflections on Ways Forward to Promote Free Expression

Given that "the tide around the world is turning against free speech,"[172] seeking to re-negotiate or amend these treaties to promote broad speech protections at the international level is inadvisable because governments would likely develop treaty text that undermines existing international speech protections. The Organization of Islamic Cooperation's attempt to develop a protocol to the CERD that would have criminalized the "defamation of religions" concept was a recent case study in which such a scenario almost played out.[173] Even European governments have sought to diminish speech protections in the international system by invoking the European Court of Human Rights' jurisprudence when defending their domestic laws before the UN's machinery.[174]

It is also inadvisable to encourage states to take reservations that lament and commemorate broad interpretations of the mandatory bans on speech that exist in the ICCPR and CERD. Instead, states that join these treaties should be encouraged to issue understandings upon ratification that reinforce the substance of the UN human rights machinery's recent interpretations, including those that apply ICCPR Article 19's tripartite test to mandatory speech bans and otherwise limit the potential scope of ICCPR Article 20 and CERD Article 4. Such understandings could propel more restrictive interpretations of ICCPR Article 20 and CERD Article 4 rather than fueling broader readings of these provisions.[175]

That said, it is unclear if the small number of countries that have not yet joined these treaties would do so with understandings that promote freedom of expression.[176] An

172. *Id.* at 55.

173. *See supra* notes 133–134 and accompanying text.

174. *See supra* note 104 and accompanying text.

175. *See supra* Part II.A–II.B.

176. *See supra* notes 138–149 and accompanying text.

important avenue for human rights advocacy therefore is to encourage and promote the UN human rights machinery's most recent interpretations in a variety of ways.[177] For example, human rights advocates should be vigilant to make sure there is not backsliding on these interpretations by the UN human rights machinery, which is constantly under governmental pressure to expand the scope of acceptable speech restrictions.[178] Advocates should engage with the treaty bodies and the UN Special Rapporteur on the range of their work (e.g., reviewing states' human rights records, considering individual complaints, and issuing interpretative guidance) and continue to reinforce the proper implementation of these interpretations in the bodies' ongoing work. If the UN Human Rights Committee or the UN CERD Committee revisit the scope of acceptable speech restrictions in future general comments or recommendations, human rights advocates should actively participate in those public consultations. In addition, human rights advocates should carefully monitor and engage during the selection processes for candidates for the UN's treaty monitoring bodies and special rapporteur positions to ensure that selected persons hold views on speech that align with the most recent UN guidance. To facilitate the migration of UN norms on speech to the regional systems, human rights advocates can highlight UN approaches in advocacy efforts at the regional level.

In essence, freedom of expression advocates should actively seek to entrench the victories that have been achieved both in the text of ICCPR Article 19 and in the most recent interpretations by the UN human rights machinery. These

177. Such interpretations include those relating to ICCPR Article 19's tripartite test, the application of Article 19's tripartite test to ICCPR Article 20 and CERD Article 4 and the understanding that "incitement" requires findings of an intent to incite, the likelihood of harm, and the imminence of that harm. *See supra* notes 26–73, 80–82 and accompanying text. These interpretations by the UN's human rights machinery are reasonable interpretations of the textual language in the ICCPR and CERD and help to realize the authors' vision for broad protections for freedom of expression, even insulting or offensive speech.

178. *See* Alberto Cerda Silva, *supra* note 127 (reporting that there is a "global assault on freedom of expression" as "governments are stretching the meaning of what's permitted" under ICCPR Article 19).

advocates should pursue an active strategy of stealing victory (broad speech protections) from the jaws of potential defeat (ICCPR Article 20 and CERD Article 4) by promoting consistent application of positive interpretations recommended by the UN machinery.

IV. Concluding Recommendations

So where does all this leave us with respect to concerns that anchoring content moderation to international human rights law risks insufficient protections for speech as well as confusion because of inconsistent norms? With respect to the international regime's scope of coverage for offensive speech, the ICCPR's rigorous tripartite test of legality, necessity, and legitimacy, which applies to all speech restrictions (including the mandatory bans in ICCPR Article 20 and CERD Article 4), goes a long way towards accomplishing the authors' vision of broad international law protections for freedom of expression. Many laws that ban such speech will fail because they are either vague or do not reflect the least intrusive means to achieve a public interest objective. In many instances, governments can deploy good governance measures that do not require banning speech. Indeed, the UN human rights machinery has explicitly and repeatedly condemned the practice of restricting speech to protect rulers, religions or royalty. ICCPR Article 19's tripartite test and recent recommendations by both UN treaty bodies and the UN Special Rapporteur[179] are extremely helpful in protecting offensive speech and paving the way to fulfill the authors' vision for broad speech protections.

What about the concern that this body of international law is afflicted by unworkable contradictions? Part of the basis of this concern seems to have been an inappropriate conflation of the UN's human rights regime with the various *regional* human rights regimes. The international system does provide different (and often greater) speech protections than the regional systems, but that does not render the international system incoherent. It simply means there are different systems

179. For a summary of such interpretations, see *supra* Part II.A.

that interpret different treaties and come out differently in their interpretations. While this is an unfortunate state of affairs for a variety of reasons, it does not mean the UN's human rights regime is incoherent or riddled with inconsistencies. To the extent the concern about inconsistency was based upon cited discrepancies within the UN machinery's interpretations, those differences are best attributed to an evolution in the thinking of UN bodies, which evinces a trend toward greater speech protections.

Having analyzed the specific arguments in *The Right to Insult in International Law*, I continue to believe that the international human rights regime, particularly ICCPR Article 19, remains a useful resource for private platforms to incorporate into their content moderation policies. But what does that mean in practical terms? How can ICCPR Article 19's tripartite test be translated into a corporate context?

As the UN Special Rapporteur explained in 2018, in order to avoid infringing on international freedom of expression rights, platforms would first need to develop speech codes that do not contain vague bans on speech.[180] The UN human rights machinery has provided ample guidance about what constitutes a vague speech ban.[181] Is it not fair to hold global platforms, which are the most powerful speech regulators in the world, to a standard that prohibits them from imposing vague speech bans?

The platforms would also need to determine whether they are choosing the "least intrusive" means when imposing speech restrictions.[182] First, the companies would need to assess if they have means at their disposal to achieve the desired goals without infringing on speech. In practical terms, they might

180. *See UN SR 2018 Report*, *supra* note 34, at ¶¶ 26, 27 (highlighting examples of platform speech codes that contain vague phrasing); *see also* Aswad, *supra* note 30, at 46–47 (assessing Twitter's hate speech rules under ICCPR Article 19's vagueness test).

181. *See supra* Part II.A.1.

182. *See UN SR 2018 Report*, *supra* note 34, at ¶ 47 (advocating for companies to properly explain how they narrowly tailor restrictions on their platforms and how they choose the least intrusive infringements on speech); Aswad, *supra* note 30, at 47–52 (discussing how Article 19's necessity test would apply to Twitter's decision making when assessing hate speech).

consider whether they can promote digital and media literacy campaigns, speak out on issues, promote dialogues and counter-narrative approaches on contentious issues, or make technical changes to their platforms. If implementing such "good governance" measures would be insufficient, they would need to analyze if they are selecting the least intrusive means for restricting speech on their continuum of options, which include adding friction to access offensive speech, demoting speech, geo-blocking, and many more avenues.[183] Finally, companies would need to assess if the selected means is likely to work or be counterproductive. This trilogy of questions reflects the due diligence that companies would need to engage in to assure their speech restrictions are "necessary." Again, is it not fair to hold the most powerful speech regulators in human history to a standard that requires refraining from infringing on individuals' speech beyond what is necessary?

The trickiest part of translating ICCPR Article 19's tripartite test into a corporate context is the "legitimacy" prong. Are companies well situated to make public interest determinations, or would their profit-seeking motives ultimately undermine such assessments? As I have argued previously, this prong requires further multi-stakeholder consultation to develop a way forward.[184] It may be that this prong should be excluded in the corporate context (which means companies would only be held to the legality and necessity conditions of the ICCPR Article 19's tripartite test). Alternatively, companies could be required to articulate which public interest goals they are pursuing with specificity and evidence-based arguments, subject to heightened scrutiny because of their brand management incentives when seeking to ban content.

---

183. *See* Aswad, *supra* note 30, at 50 (analyzing some of the options available to platforms to restrict speech, including deleting specific posts but still allowing the user to utilize the platform and issuing warnings to users who continuously violate "a company's speech code").

184. *See id.* at 52–57 ("Applying the third prong of Article 19(3)'s tripartite test raises a number of questions that would benefit from further conversations by interested stakeholders to assess the contours of what is feasible and to avoid corporations invoking public interest rationales as pretexts for revenue-driven decisions.").

We are at a liminal moment in which global social media companies are deciding whether to align their worldwide speech moderation with international human rights law. This international regime provides a useful framework and significant guidance that these companies can deploy in the content moderation context, particularly with respect to ICCPR Article 19's legality and necessity conditions. The existing international regime that governs freedom of expression has the benefit of reigning in the vast discretion of platforms in content moderation through the use of a principled framework that has the added benefit of legitimacy rooted in global standards. We should be encouraging companies to voluntarily align their content moderation policies with this international regime.

In other words, this is another opportunity to steal victory (i.e., broad international protections for freedom of expression for billions) from the jaws of potential defeat (i.e., vague corporate speech bans that can be imposed without regard to workable measures that infringe less on expression). So why not steal victory from the jaws of defeat?