

7-7-2022

Convergence & Conflict: Reflections on Global and Regional Human Rights Standards on Hate Speech

Evelyn Aswad

David Kaye

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/njihr>



Part of the [Human Rights Law Commons](#), and the [International Law Commons](#)

Recommended Citation

Evelyn Aswad and David Kaye, *Convergence & Conflict: Reflections on Global and Regional Human Rights Standards on Hate Speech*, 20 NW. J. HUM. RTS. 165 (2022).

<https://scholarlycommons.law.northwestern.edu/njihr/vol20/iss3/1>

This Article is brought to you for free and open access by Northwestern Pritzker School of Law Scholarly Commons. It has been accepted for inclusion in Northwestern Journal of Human Rights by an authorized editor of Northwestern Pritzker School of Law Scholarly Commons.

CONVERGENCE & CONFLICT: REFLECTIONS ON GLOBAL AND REGIONAL HUMAN RIGHTS STANDARDS ON HATE SPEECH

*Evelyn Aswad** & David Kaye*¹*

ABSTRACT—What is hate speech under international human rights law? And how do key international adjudicators interpret the law governing it? This Article seeks to illuminate two countervailing and under-reported trends: on the one hand, a growing consensus among U.N. experts and treaty bodies concerning interpretations of “hate speech” prohibitions in international law; and on the other, a failure of several regional human rights bodies to develop approaches to hate speech that are consistent with the U.N.’s universal standards. The Article begins by analyzing the U.N.’s approach to freedom of expression and hate speech and examining how, in the last decade, various U.N. expert bodies have converged on an agreed approach to the subject. The Article next compares this global standard with key developments in the Inter-American, European, and African human rights systems and the emerging frameworks in Arab, Islamic, and Southeast Asian contexts. This comparative analysis reveals that, while certain systems converge with the U.N.’s approach, others diverge, sometimes marginally, sometimes significantly. For example, the European Court of Human Rights frequently lessens or removes the burden on governments to show hate speech restrictions are properly imposed, allows for the imposition of hate speech restrictions for reasons not accepted at the global level, and does not assess whether restrictions on speech are the least intrusive means to achieving legitimate public interest objectives. After analyzing this landscape of regional norms in convergence and conflict with U.N. standards, the Article provides several observations. The Article concludes by urging human rights defenders throughout the world to be cognizant of the areas in which regional human rights bodies provide fewer

¹ **Herman G. Kaiser Chair in International Law and Director of the Center for International Business & Human Rights at the University of Oklahoma College of Law. The author would like to thank her research assistants for their outstanding work: Robert Rembert, Morgan Vastag, Taryn Chubb, and Maddison Craig.

*Clinical Professor of Law and Director of the International Justice Clinic, University of California (Irvine) School of Law; Former United Nations Special Rapporteur on the Right to Freedom of Opinion and Expression (2014-2020).

protections than U.N. standards require, and to tackle this trend through proposed strategies to protect universal minimum standards for freedom of expression.

TABLE OF CONTENTS

TABLE OF CONTENTS..... 166
 INTRODUCTION 166
 I. INTERNATIONAL STANDARDS ON FREEDOM OF EXPRESSION
 AND HATE SPEECH 172
 A. Protection of Speech and Permissible Restrictions 172
 B. The U.N.'s Mandatory Bans on Certain Hate Speech 176
 C. Application of Tripartite Test to Hate Speech Bans 182
 II. REGIONAL HUMAN RIGHTS STANDARDS ON EXPRESSION AND
 HATE SPEECH..... 186
 A. The Inter-American Human Rights System 186
 B. The European Human Rights System 190
 C. The African Human Rights System 198
 D. Emergent Regional Approaches 202
 III. REFLECTIONS & RECOMMENDATIONS 206
 A. Comparison of Global and Regional Standards 206
 B. Recommendations for Human Rights Defenders 209
 CONCLUSION 214

INTRODUCTION

The summer of 2015 was an unusual one in Germany, tense and transformative. Chancellor Angela Merkel had declared that Germany would accept up to one million refugees, the vast majority of whom were fleeing civil war in Syria.² In turn, right-wing agitation and violence increased against refugees and minorities, including those long in residence, as well as citizens of Germany.³ From the perspective of the German government, the spread of racism and hate speech on online platforms, especially Facebook, contributed to discrimination and violence.⁴ Heiko Maas, then the German Minister of Justice, demanded that Facebook take action against hate speech on its platform.⁵ The demand set in motion a series of governmental efforts, first to create industry codes of

² DAVID KAYE, SPEECH POLICE: THE GLOBAL STRUGGLE TO GOVERN THE INTERNET 65 (2019).

³ *Id.* at 66.

⁴ *Id.* at 66-67.

⁵ *Id.*

conduct against hate speech in Europe,⁶ and ultimately to adopt a German law that required the largest platforms to take down hate speech and other content deemed illegal under domestic law – a controversial statute known as the Network Enforcement Act, or NetzDG.⁷

This debate triggered basic questions: what *is* hate speech? How should it be defined and adjudicated? What type of expression constitutes hate speech that should be subject to takedowns by companies and prohibition by the state? Should the law vary according to whether speech is online or offline? What boundaries on such laws are imposed by human rights treaties? Even online platforms themselves have turned to international human rights law to try to answer these knotty questions of content.⁸ Indeed, the need for clarity on hate speech law is evident as it is a topic of global attention, national regulation, political debate, and social media anxiety.⁹

⁶ *Id.* See also Evelyn Aswad, *The Role of U.S. Technology Companies as Enforcers of Europe's New Internet Hate Speech Ban*, 1 COLUM. HUM. RTS. L. REV. ONLINE 1, 3-4 (2016) (describing the main provisions of Europe's Internet hate speech code of conduct for companies).

⁷ KAYE, *supra* note 2, at 67-70. NetzDG required large platforms to “remove content that violated provisions of German law” within short time frames or face hefty financial penalties. *Id.* at 68-69. Human rights watchdogs and companies objected to the law because, *inter alia*, it inappropriately privatized the role of courts and public prosecutors with respect to speech adjudication matters and created problematic incentives for companies to remove speech without “any countervailing pressure to keep legitimate content up.” *Id.* at 69-70. Typically, hate speech adjudications by public authorities require much more time consuming and careful review and deliberation. See JACOB MCHANGAMA ET AL., FUTURE OF FREE SPEECH PROJECT, RUSHING TO JUDGMENT: ARE SHORT MANDATORY TAKEDOWN LIMITS FOR ONLINE HATE SPEECH COMPATIBLE WITH THE FREEDOM OF EXPRESSION? 17-30 (2021), https://futurefreespeech.com/wp-content/uploads/2021/01/FFS_Rushing-to-Judgment-3.pdf (comparing state-mandated timeframes for platforms to remove online hate speech with the average time that public authorities utilize in speech adjudication proceedings). Eventually, the NetzDG law served as a template for authoritarian regimes to suppress online speech. See Jacob Mchangama & Joelle Fiss, *Germany's Online Crackdowns Inspire the World's Dictators*, FOREIGN POL'Y (Nov. 6, 2019), <https://foreignpolicy.com/2019/11/06/germany-online-crackdowns-inspired-the-worlds-dictators-russia-venezuela-india>.

⁸ See, e.g., FACEBOOK [META], CORPORATE HUMAN RIGHTS POLICY (2021), <https://about.fb.com/wp-content/uploads/2021/04/Facebooks-Corporate-Human-Rights-Policy.pdf> (noting that Facebook is “committed to respecting human rights as set out in the United Nations Guiding Principles on Business and Human Rights.”) [hereinafter FACEBOOK, CORPORATE HUMAN RIGHTS POLICY]; Jack Dorsey (@jack), TWITTER (Aug. 10, 2018, 9:58 AM), <https://twitter.com/jack/status/1027962500438843397> [<https://perma.cc/A297-PPMA>] (setting forth the view of Twitter's CEO that his company should ground its values in international human rights law).

⁹ See, e.g., *United Nations Strategy and Plan of Action on Hate Speech*, UNITED NATIONS, <https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml> (2019) (noting the U.N. Secretary General launched a strategic plan to tackle “alarming trends” in the growth of hate speech and intolerance); Gilad Edelman, *The Parler Bans Open a New Front in the 'Free Speech' Wars*, WIRED (Jan. 13, 2021, 7:00 AM), <https://www.wired.com/story/parler-bans-new-chapter-free-speech-wars> (observing that the ban on Parler for failing to tackle hate speech “opens a new front in the online speech wars”); Mathieu Rosemain, *Twitter Loses Online Hate Speech Court Battle in France*, REUTERS (Jan. 20, 2022, 8:26 AM), <https://www.reuters.com/world/europe/twitter-loses-appeal-french-case-over->

International human rights law standards have converged around certain common approaches to hate speech. Consider, for instance, the key interpretations of the U.N. Human Rights Committee, which monitors implementation of the International Covenant on Civil and Political Rights (“ICCPR”),¹⁰ and the U.N. Committee on the Elimination of Racial Discrimination, which monitors implementation of the International Convention on the Elimination of All Forms of Racial Discrimination (“ICERD”).¹¹ Both treaty bodies have found that hate speech prohibitions must adhere to a test that, *inter alia*, forbids vague speech bans and requires that restrictions reflect the least intrusive means to further legitimate public interest objectives.¹² These U.N. treaty bodies, along with expert opinions offered within the U.N. system, have developed an increasingly consistent set of rules governing the appropriate boundaries for hate speech laws.¹³

Regional human rights institutions have not always followed the lead of the U.N. interpretations. Some nations even seek at times to justify their violations of U.N. treaty obligations on freedom of expression by invoking *regional* human rights norms rather than adhering to the *global* standards that are also binding on them. For example, the European human rights system, though generally reinforcing global human rights, departs from international standards in ways that implicate hate speech law.¹⁴

online-hate-speech-source-2022-01-20 (reporting that a French court has required Twitter to reveal how it addresses hate speech on its platform and observing this decision “provides ammunition to campaigners elsewhere in Europe who want tougher controls to prevent the spread” of hate speech online).

¹⁰ International Covenant on Civil and Political Rights, art. 28, *opened for signature* Dec. 16, 1966, S. Exec. Doc. E, 95-2, at 29 (1978), 999 U.N.T.S. 171, 178 (entered into force Mar. 23, 1976) [hereinafter ICCPR]. The ICCPR established the Human Rights Committee, which is composed of independent experts elected by State Parties and provides guidance by (1) producing recommended interpretations of various provisions (known as General Comments), (2) reviewing periodic reports of State Parties and publishing its observations and conclusions on implementation matters, and (3) issuing decisions with respect to individual complaints when State Parties have consented to this procedure. *Id.* arts. 28-40. *See also* Optional Protocol to the International Covenant on Civil and Political Rights, *opened for signature* Dec. 16, 1966, art. 1, 999 U.N.T.S. 171, 302 (entered into force Mar. 23, 1976) (describing the Committee’s competency over individual complaints for states that have ratified the Protocol to the ICCPR).

¹¹ International Convention on the Elimination of All Forms of Racial Discrimination, arts. 8-13, *opened for signature* Dec. 21, 1965, S. Exec. Doc. C. 95-2 (1978), 660 U.N.T.S. 195 (entered into force Jan. 4, 1969) [hereinafter ICERD]. ICERD created a Committee on the Elimination of Racial Discrimination that produces guidance to the treaty’s State Parties by issuing (1) recommended interpretations of the treaty (known as General Recommendations), (2) periodic observations to State Parties about their implementation of the treaty, and (3) decisions in individual complaints against State Parties that have recognized the Committee’s competence in this regard. *Id.* arts. 8-15.

¹² *See infra* Part I(C) (analyzing the scope and application of this principled test).

¹³ *See infra* Part I(C).

¹⁴ *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*

Germany's NetzDG law reinforced its national hate speech laws that prohibit denigration of religions and the denial of historic atrocities – speech that is protected in the U.N. system but that European institutions have found can be criminalized under regional standards.¹⁵ When challenged in a U.N. inquiry, Germany defended its legislation, in part, by invoking European rather than U.N. human rights standards.¹⁶

In the same vein, the Organization of Islamic Cooperation (OIC), an inter-governmental body with 57 member states,¹⁷ recently updated its Cairo Declaration on Human Rights in Islam,¹⁸ and justified its departures from U.N. standards by citing to European case law. According to the updated Declaration, speech that offends religious sensibilities or may display religious intolerance should not be protected speech,¹⁹ conflicting

to Freedom of Opinion and Expression, ¶¶ 26-27, U.N. Doc. A/74/486 (Oct. 9, 2019), [hereinafter *Special Rapporteur 2019 Report to UNGA*] (observing that the European Court of Human Rights departs from U.N. standards by, among other things, permitting bans on blasphemy and genocide denial, determining frequently that hate speech cases are inadmissible as an abuse of freedom of expression, and deferring significantly to governments). See also *infra* notes 236-58 and accompanying text for a summary of key divergences between the U.N. and European systems with respect to hate speech.

¹⁵ Evelyn Aswad, *Are Recent Governmental Initiatives to Combat Online Hate Speech, Extremism, and Fraudulent News Consistent with the International Human Rights Law Regime?* in GOVERNANCE INNOVATION FOR A CONNECTED WORLD: PROTECTING FREE EXPRESSION, DIVERSITY AND CIVIC ENGAGEMENT IN THE GLOBAL DIGITAL ECOSYSTEM 29, 30-33 (2018), <https://www.cigionline.org/sites/default/files/documents/Stanford%20Special%20Report%20web.pdf> (noting the NetzDG law requires social media companies take action against a range of speech criminalized under German law, including denigration of religious faiths and denial of facts related to the Nazi era, and comparing such bans to U.N. and European human rights standards).

¹⁶ See Letter from the 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) 2-3, (June 1, 2017), <https://perma.cc/95YG-MKAK>. Specifically, the German government's response implied that protections under the European Convention on Human Rights would be sufficient to meet its freedom of expression obligations under U.N. treaties. See *id.* at 2 (“Since the German Constitution and the European Convention on Human Rights protect freedom of expression, there will also be effective remedies for anyone wishing to challenge any actions based on the new law.”). In addition, the German government invoked the European Court of Human Rights’ jurisprudence, which allows for broad bans on hate speech, in justifying its new law. *Id.* at 3 (“Likewise, the European Court of Human Rights in its case-law has made abundantly clear that hate speech is intolerable in a democratic society.”).

¹⁷ *History*, OIC, https://www.oic-oci.org/page/?p_id=52&p_ref=26&lan=en (last visited Feb. 20, 2022) [hereinafter *OIC History*] (commemorating the organization’s membership and noting that it is the second largest intergovernmental organization after the United Nations).

¹⁸ The Cairo Declaration on Human Rights in Islam, *adopted* Nov. 28, 2020, OIC-IPHRC, http://oic-iphrc.org/ckfinder/userfiles/files/FINAL%20OHRD%20CLEAN%20%20VERSION%2024_12_2020.pdf [hereinafter *Cairo Declaration*].

¹⁹ *Id.* art. 21(c) (“Freedom of expression should not be used for denigration of religions and prophets or to violate the sanctities of religious symbols”). Prior to the adoption of the revised declaration, the OIC’s Human Rights Commission had made clear that “sheer disrespect, defamation, insult, and negative stereotyping” constitute “incite[ment of] religious hatred” and would not be protected speech. *Indep.*

with U.N. standards that protect speech such as blasphemy.²⁰ The OIC's Independent Permanent Human Rights Commission (OIC-IPHRC) praised the European Court of Human Rights' reluctance to prohibit the criminalization of blasphemy, or speech that offends religious sensibilities or displays religious intolerance.²¹

Such instances of resorting to regional jurisprudence to rationalize providing fewer human rights protections than those embodied in global standards are problematic for a variety of reasons. First, as a matter of established treaty law, "*regional* human rights norms cannot, in any event, be invoked to justify departure from *international* human rights protections."²² In other words, violating the minimum standards in U.N. human rights treaties cannot be legally justified by citing to differing regional norms. Second, the divergence between global and regional norms on hate speech often causes inappropriate conflation of, and confusion about, the relevant standards, which can lead to concerns that international

Permanent Hum. Rts. Comm. [IPHRC], 8th Session: Outcome Document of the Thematic Debate on 'Freedom of Expression' and 'Hate Speech', at 1 (Nov. 23, 2015), https://oic-iphrc.org/en/data/docs/session_reports/8th/8th_iphrc_thematic_debate_outcome_en.pdf.

²⁰ See *infra* notes 103-05 and accompanying text (describing the U.N. Human Rights Committee's approach to blasphemy).

²¹ *On the Occasion of 'International Human Rights Day' IPHRC Welcomes the Adoption of 'Cairo Declaration of the OIC on Human Rights' and Calls upon Member States to Apply Human Rights Based Approaches in Dealing with the Socio-Economic and Health Related Challenges of COVID-19 Pandemic*, OIC-IPHRC, (Dec. 10, 2020), <https://oic-iphrc.org/home/article/479> ("the Commission recalled and reaffirmed . . . the jurisprudence emerging from the European Court of Human Rights, which validates restrictions on freedom of expression criticizing religious beliefs where such expression constitutes incitement to hatred and is deemed offensive to the adherents of a particular religion."). The European Court has found that restricting speech to avoid offending religious sensibilities is a legitimate governmental objective. See *infra* note 172 and accompanying text.

²² *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 26 (emphasis added). Under the rules for treaty interpretation that are set forth in the Vienna Convention on the Law of Treaties, regional human rights treaties would not qualify as sources that define the scope of global treaties. See Vienna Convention on the Law of Treaties art. 31(2)-(3), *opened for signature* May 23, 1969, 1155 U.N.T.S. 331, 340 (entered into force Jan. 27, 1980) (noting that only treaties that all parties to original treaty have endorsed as defining the scope of the original treaty are appropriate sources of interpretation). Given that *regional* human rights treaties have not been endorsed by all the parties to *U.N.* human rights treaties, such *regional* treaties do not define the scope of *global* treaty obligations. Indeed, it would undermine the point of global treaties and certainty in multilateral treaty relations if a regional group could develop its own treaty that changed the scope of a global treaty's obligations. If resort is made to regional treaties and their relevant jurisprudence to learn lessons when examining the scope of global treaties, great care should be taken when making such comparative assessments. See Gerald Neuman, *The Draft General Comment on Freedom of Assembly: Might Less Be More*, JUST SECURITY (Feb. 4, 2020), <https://www.justsecurity.org/68465/the-draft-general-comment-on-freedom-of-assembly-might-less-be-more/> (urging the U.N. Human Rights Committee "to demonstrate its commitment to conscientiously interpreting its own global Covenant" rather than relying on regional jurisprudence without examination of relevant differences between global and regional approaches and treaties).

law provides little guidance to states or online platforms seeking to regulate hateful speech.²³ Third, these normative divergences on hate speech allow governments to couch their challenges to universal norms by citing to regional jurisprudence, rather than more openly acknowledging that their challenge is based on domestic law approaches. In essence, the use of regional jurisprudence to justify violating U.N. norms may at times be a contemporary manifestation of a persistent challenge to the universal human rights project: the invocation of cultural relativism to legitimize the failure to implement global minimum standards.

This cocktail of considerations creates conditions that can eventually erode global human rights standards, which makes this a liminal moment to examine how U.N. and regional standards converge and conflict with regard to hate speech, to reflect on the potential impacts of normative divergences, and to develop ways forward. Part I of this Article assesses how key components of the U.N.'s human rights machinery have converged in their approach to both discretionary and mandatory hate speech bans. Part II examines how regional human rights systems in Europe, the Americas, and Africa approach freedom of expression and hate speech. This Part also notes emergent conflicts with normative frameworks in the new and weakly-institutionalized approaches of the Arab region, the Association of Southeast Asian Nations, and the Organization of Islamic Cooperation. Part III begins by extrapolating from this mapping exercise how regional approaches converge and conflict with U.N. standards. It highlights that the European human rights system in particular departs from the U.N.'s minimum protections for freedom of expression, which not only challenges the U.N. system but also can (adversely) influence the development of other regional freedom of expression norms. This Part next provides recommendations about how human rights defenders can reinforce U.N. approaches to hate speech, which are grounded in principled measures for both protecting speech and tackling intolerance.²⁴

²³ See Evelyn Mary Aswad, *To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?*, 77 WASH. & LEE L. REV. 609, 614-43 (2020) (describing how scholars and lawyers have conflated global and regional norms to conclude that international law provides conflicting and insufficient guidance on freedom of expression issues, including hate speech).

²⁴ Our reference to "human rights defenders" intends to be addressed not only to national and regional civil society actors but also lawyers, independent experts in the regional and U.N. human rights environments, and others working to defend freedom of expression norms and to protect individuals and communities against the discrimination, hostility, and violence that ICCPR Article 20(2) seeks to address. See *infra* notes 56-74 and accompanying text for an analysis of ICCPR Article 20. Hate speech prohibitions that go beyond those authorized by U.N. standards have not been shown to effectively promote human rights, including with respect to equality and non-discrimination, but rather often undermine them. See generally NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE

I. INTERNATIONAL STANDARDS ON FREEDOM OF EXPRESSION AND HATE SPEECH

The freedom of expression as a human right under international law centers on two foundational instruments established under the umbrella of the United Nations: the 1948 Universal Declaration of Human Rights (UDHR) and the 1966 International Covenant on Civil and Political Rights (ICCPR). The UDHR provides, “Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.”²⁵ Like all the rights in the UDHR, the freedom of expression may be restricted under certain limited circumstances.²⁶ The ICCPR refined and codified this standard.²⁷ Part I analyzes U.N. standards on freedom of expression and hate speech, focusing on the ICCPR and the authoritative interpretations that have developed around it. It begins by examining the ICCPR’s standard of broad protection for freedom of expression as well as its requirement that any speech restrictions comply with a three-part test of legality, necessity and legitimacy. It next analyzes the scope of mandatory hate speech bans under the ICCPR and ICERD as well as how these mandatory bans intersect with the ICCPR’s test for speech restrictions. Part I concludes that key human rights bodies within the U.N. system have largely converged in their approach to hate speech.

A. *Protection of Speech and Permissible Restrictions*

Article 19(2) of the ICCPR provides: “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.”²⁸ This broad protection for freedom of expression explicitly covers both the ability to “impart” and to “receive” all kinds and all forms of expression, encompassing the rights of speakers and audiences. ICCPR

SPEECH, NOT CENSORSHIP 133-82 (2018) (explaining how censorship is ineffective and often counterproductive in tackling intolerance and why non-censorial methods are more effective in combating hate and intolerance).

²⁵ G.A. Res. 217 (III) A, Universal Declaration of Human Rights, art. 19 (Dec. 10, 1948) [hereinafter UDHR].

²⁶ Any limitation on UDHR rights must be “determined by law solely for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society.” *Id.* art. 29(2).

²⁷ ICCPR, *supra* note 10, art. 19. See Part I(A) for a discussion of ICCPR art. 19.

²⁸ ICCPR, *supra* note 10, art. 19(2) (emphasis added).

Article 19(3) acknowledges that limitations on expression may be permitted, but subjects them to a strict three-part test.²⁹ The government bears the burden of proving that any restriction on expression is (1) provided by law and (2) necessary (3) for a legitimate public interest objective.³⁰ This tripartite test is often referred to as the cumulative conditions of legality, necessity, and legitimacy.³¹ The U.N. human rights machinery has made clear that states should apply this test both when adopting legislation and applying laws that restrict speech to particular situations.³²

In 2011, after multi-year deliberations, the Human Rights Committee produced interpretive guidance on Article 19, in General Comment 34.³³ Its lead drafter noted that General Comment 34 reflects an evolution of the Committee's approach by broadening its interpretations of the scope of freedom of expression and providing "in the most elaborated fashion" to date its views on the tripartite test.³⁴ General Comment 34 emphasizes that this three-part test applies to *all* state restrictions on speech³⁵ and that states bear the burden of demonstrating their lawfulness.³⁶ Moreover, states "must demonstrate in a specific and individualized fashion the precise nature of the threat . . . [including] by establishing a direct and immediate connection between the expression and the threat."³⁷

²⁹ *Id.* art. 19(3) (providing that the government "may" rather than "must" restrict speech).

³⁰ *Id.* art. 19(3); U.N. Human Rights Committee, General Comment No. 34, ¶¶ 27, 35, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011).

³¹ 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*, ¶ 11, U.N. Doc. A/71/373 (Sept. 6, 2016) [hereinafter *Special Rapporteur 2016 Report to UNGA*].

³² *See, e.g.*, Irene Khan et al., Special Rapporteur on the Promotion and Protection of the Freedom of Opinion and Expression, Comments on the Intermediary Guidelines and Digital Media Ethics Code, 2021, OL IND 8/2021 (June 11, 2021), <https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gId=26385> (expressing concern about the violation of freedom of expression standards with regard to legal regulations); Irene Khan et al., Special Rapporteur on the Promotion and Protection of the Freedom of Opinion and Expression Joint Allegation Letter to Bahrain, AL BHR 3/2021 (Sept. 16, 2021), <https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gId=26630> (expressing concern about the violation of freedom of expression obligations with respect to the treatment of a particular individual).

³³ 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, 645-53 (2012) (describing the process that was used to adopt General Comment No. 34).

³⁴ *Id.* at 647-53.

³⁵ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶¶ 50-52.

³⁶ *Id.* ¶ 27.

³⁷ *Id.* ¶ 35.

1. *Legality*

The Human Rights Committee has stated that any restriction on speech, in order to be “provided by law,” must not be improperly vague.³⁸ Restrictions on speech must give appropriate notice to the public of what speech is banned while appropriately constraining the discretion of governmental authorities.³⁹ Further, any law restricting speech must be properly enacted within a country’s domestic legal system.⁴⁰ The legality prong also encompasses a ban on overly broad restrictions (i.e., restrictions that cover more speech than needed), requires that restrictions be subject to public comment, and implies that independent judicial officials must oversee the implementation of speech restrictions.⁴¹

2. *Legitimacy*

Any restriction on expression must be designed to protect a legitimate, enumerated objective set forth in ICCPR Article 19(3): the protection of the rights or reputations of others, national security, public order, public health or morals.⁴² The Human Rights Committee has sought to constrain these potentially open-ended grounds for speech restrictions. For instance, the Committee explained that the purpose of protecting public morals should not be based exclusively on a “single tradition” and must be interpreted “in the light of the universality of human rights” and non-discrimination principles.⁴³ It has criticized governmental invocations of public interest objectives that serve as pretexts for illegitimate interests.⁴⁴ The Committee has emphasized that governments may not invoke additional grounds beyond those enumerated in Article 19(3) to justify speech restrictions.⁴⁵

³⁸ *Id.* ¶ 25.

³⁹ *Id.*

⁴⁰ *Id.* ¶ 24; *Special Rapporteur 2016 Report to UNGA*, *supra* note 31, ¶ 12.

⁴¹ *Special Rapporteur 2016 Report to UNGA*, *supra* note 31, ¶¶ 12, 13; 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*, U.N. Doc. A/HRC/38/35, ¶ 7 (Apr. 6, 2018); *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 6(a).

⁴² ICCPR, *supra* note 10, art. 19(3).

⁴³ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶ 32. U.N. human rights monitors have implemented this guidance with respect to various laws in different regions. *See, e.g.*, Letter of U.N. Special Rapporteurs to the Government of Lebanon, AL LBN 3/2019 (May 10, 2019), <https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gId=24560> (criticizing government blocking of the largely gay dating app, Grindr as an improper invocation of morals); Letter of U.N. Special Rapporteurs to the Government of Ghana, OL GHA 3/2021 (Aug. 9, 2021), <https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gId=26586> (criticizing Ghanaian legislation seeking to limit activities of individuals on the basis of their sexual orientation as “not in keeping with the proper definition of ‘public morals.’”).

⁴⁴ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶ 30.

⁴⁵ *Id.* ¶ 22.

3. Necessity

Limitations on speech must also, among other things, be (1) the least intrusive means to achieve the legitimate objective and (2) proportionate to the interest to be protected.⁴⁶ A three-prong approach can assist in assessing whether a speech restriction qualifies as the least intrusive means to achieve a legitimate objective.⁴⁷ First, the state should assess if it has or can create any non-censorial methods to achieve the goal. For example, governments have the ability to engage in a variety of measures to combat intolerance that do not burden speech. They can, *inter alia*, implement laws prohibiting acts of discrimination, deploy high level governmental officials to speak out against intolerance, and train governmental officials to prevent discriminatory profiling and other practices. If such avenues are sufficient to combat intolerance, then restricting speech is not the least intrusive means.⁴⁸ Second, if methods that do not burden speech are insufficient, then the state should rank its available speech-restricting tools and select the one that least burdens freedom of expression, paying particular attention to the right of the speaker to speak and of listeners to hear the speaker's views.⁴⁹ This involves assessing, among other things, a range of civil and criminal penalties and demonstrating selection of the least intrusive means. Third, the state should also demonstrate through transparent, evidence-based reasoning that the restriction serves to achieve the legitimate objective.⁵⁰ If restrictions are ineffective or counterproductive, then they cannot meet the test of necessity.⁵¹

⁴⁶ *Id.* ¶ 34.

⁴⁷ *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 52.

⁴⁸ *See id.*

⁴⁹ *See id.*

⁵⁰ *See id.* U.N. monitors have frequently asserted that states should provide justifications that are evidence-based when restricting speech. *See, e.g.*, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Third Party Intervention filed in the European Court of Human Rights*, ¶ 20 (June 3, 2019), Microsoft Word - Big Brother Watch draft DK final look.docx (ohchr.org), ("States regularly invoke national security to legitimise surveillance measures that entail over-broad restrictions on human rights. The invocation of national security does not in and of itself provide an adequate human rights law justification. Rather, the State must provide an 'articulable and evidence based justification for the interference'. The State must, at a minimum, give a meaningful public account of the tangible benefits."); David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Research Paper 1/2019 on Freedom of Expression and Elections in the Digital Age*, at 11 (June 2019), <https://www.ohchr.org/Documents/Issues/Opinion/ElectionsReportDigitalAge.pdf> ("approaches for combating disinformation should be evidence-based and tailored to proven or documented impacts of disinformation or propaganda."); Joint Declaration of Freedom of Expression Mandate Holders on Freedom of Expression and Countering Violent Extremism (May 4, 2016), <https://www.osce.org/files/f/documents/e/9/237966.pdf> ("... [h]ighlighting that . . . programmes and initiatives that restrict freedom of expression must be based on evidence of their effectiveness . . .").

⁵¹ *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 52.

In keeping with the least-intrusive means standard, the U.N. system has been working with governments to generate approaches to counter hatred and intolerance that avoid restrictions on speech. For example, U.N. Human Rights Council Resolution 16/18 achieved consensus on an approach to tackling religious intolerance that focuses on governments exercising non-censorial means as a first resort and deploying bans on speech only when there is “incitement to imminent violence.”⁵² Such non-censorial means include governments legislating against religious discrimination and hate crimes, engaging in inter-faith dialogues, promoting education on tolerance, reaching out to vulnerable populations to assist with grievances, and having governmental officials speak out against intolerance.⁵³ Resolution 16/18 underscored that a human rights-based approach to tackling hate and intolerance means governments generally should not resort to speech bans before attempting non-restrictive measures.

The proportionality test, as part of necessity, requires governments to demonstrate that restrictions “target a specific objective and [do] not unduly intrude upon the other rights of targeted persons.”⁵⁴ In addition, the proportionality assessment also means states restricting speech should show “[t]he ensuing ‘interference with third parties’ rights [is] limited and justified in the light of the interest supported by the intrusion.”⁵⁵ It should be noted that, while they have aspects that may intersect, the least intrusive means test and the proportionality test are not the same. The government must meet its burden under both tests for a speech restriction to survive scrutiny. The proportionality analysis operates as a double-check on governments to demonstrate the necessity of speech restrictions by showing that the benefit to the public interest objective outweighs the burden on speech and other interests, even if the least intrusive means test is met.

B. The U.N.’s Mandatory Bans on Certain Hate Speech

Mandatory hate speech bans in U.N. human rights treaties are the principal exception to the normal rule of Article 19(2) guaranteeing freedom to seek, receive and impart “information and ideas of all kinds.”

⁵² Human Rights Council Res. 16/18, U.N. Doc. A/HRC/RES/16/18, ¶ 5(f) (Apr. 12, 2011) [hereinafter Council Res. 16/18].

⁵³ *Id.* ¶ 5(a), (b), (e).

⁵⁴ David Kaye (Special Rapporteur on the Right to Freedom of Opinion and Expression) et al., *Mandates of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; and the Special Rapporteur on Freedom of Religion or Belief*, Comm. No. OL USA 6/2017, 3 (May 9, 2017), www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-USA-6-2017.pdf.

⁵⁵ *Id.*

This section begins by examining ICCPR Article 20's mandatory ban on advocacy of hatred that rises to the level of incitement to discrimination, hostility or violence. It next examines ICERD's ban on racist hate speech. This section addresses how each of these mandatory bans intersects with ICCPR Article 19(3)'s tripartite test, which applies to all speech restrictions.

1. *ICCPR Article 20(2)*

Article 20(2) requires the prohibition of “any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.”⁵⁶ In 2006, the Human Rights Council requested that the U.N. Office for the High Commissioner for Human Rights produce a report evaluating the law on “incitement to racial and religious hatred . . . [and] its implications for” ICCPR Article 20(2).⁵⁷ The High Commissioner's report assessed that there was no consensus among states about the scope of keywords in Article 20.⁵⁸ In 2011, after the adoption of both Resolution 16/18 and the Human Rights Committee's General Comment 34, U.N. human rights mechanisms sought to delineate the appropriate scope of ICCPR Article 20.⁵⁹

In September 2012, the U.N. Special Rapporteur on freedom of opinion and expression – appointed by the Human Rights Council to monitor state compliance with Article 19 and other free speech norms in international human rights law – noted that a state must demonstrate three threshold elements for speech to qualify for prohibition under ICCPR Article 20: intent, incitement, and particular harms.⁶⁰ First, there must be

⁵⁶ ICCPR, *supra* note 10, art. 20.

⁵⁷ H.R.C. Dec. 1/107, U.N. Doc. A/HRC/DEC/1/107 (June 30, 2006).

⁵⁸ U.N. High Comm'r for Hum. Rts., Rep. on Incitement to Racial and Religious Hatred and the Promotion of Tolerance, U.N. Doc. A/HRC/2/6, ¶ 3 (Sept. 20, 2006).

⁵⁹ It should be noted that the Human Rights Committee has not yet provided comprehensive views on the substantive scope of ICCPR Article 20. In 1983, it issued a General Comment that focused primarily on improved reporting in States Parties' periodic reports. U.N. Human Rights Committee, General Comment No. 10, art. 19 (Freedom of Opinion and Expression), 19th Sess. (adopted June 29, 1983). In General Comment 11, adopted one month later, the Human Rights Committee mainly restated the prohibitions required under Article 20 and urged states to implement them by law. U.N. Human Rights Committee, General Comment No. 11, art. 20 (Prohibition of propaganda for war and inciting national, racial or religious hatred), 19th Sess. (adopted July 29, 1983). In 2017, in a decision rendered as part of its individual complaints process, the Committee found no violation of ICCPR Article 20(2) when a State Party had adequate laws and procedures to implement this article and had prosecuted but did not convict a particular speaker. *Rabbae v. Netherlands* (Hum. Rts. Comm. Commc'n No. 2124/2011), ¶ 10.7 (Mar. 29, 2017).

⁶⁰ See, e.g., Frank La Rue (Special Rapporteur on 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*, U.N. Doc. A/67/357, ¶ 43 (Sept. 7, 2012) [hereinafter *Special Rapporteur 2012 Report to UNGA*].

“advocacy” of national, racial or religious hatred, which means the speaker has a specific intent to promote hatred towards the target group and does so publicly.⁶¹ Second, the speech must rise to “incitement” against a targeted group, which means that the statements should “create an imminent risk of discrimination, hostility or violence against persons belonging to those groups.”⁶² In this regard, he noted that “advocacy of hatred . . . is not an offence in itself” but rather the speech must rise to the level of incitement to the harms listed in ICCPR Article 20.⁶³ Third, therefore, the incitement must be directed to producing one of three harmful results: discrimination, hostility, or violence against the targeted group.⁶⁴ In acknowledging the potential ambiguity (and lack of jurisprudence) around “hostility,” the Special Rapporteur highlighted that this term should be interpreted to mean “a manifestation of hatred beyond a mere state of mind.”⁶⁵

The Special Rapporteur also observed that Article 20 requires “prohibition” of speech that meets its criteria, a reference to civil sanctions.⁶⁶ *Criminal* sanctions for incitement under Article 20, the Special Rapporteur found, should be reserved for the most serious and extreme cases, such as those involving imminent violence.⁶⁷ In 2011, the Human Rights Committee stated that restrictions imposed under ICCPR Article 20 must also meet Article 19’s tripartite test.⁶⁸ The Special Rapporteur has emphasized that ICCPR Article 20 protects the target of a speaker’s advocacy of hatred and does not address (or in any way condone or permit) a target group’s reaction to ‘offensive’ speech with violence.⁶⁹ There is no “heckler’s veto” in international human rights law.⁷⁰

In 2011-2012, the U.N. High Commissioner for Human Rights convened experts from around the world to develop a proposal for

⁶¹ *Id.* ¶¶ 43-44.

⁶² *Id.*

⁶³ *Id.*

⁶⁴ *Id.*

⁶⁵ *Id.* ¶ 44.

⁶⁶ *Id.* ¶ 47.

⁶⁷ *Id.*

⁶⁸ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶¶ 50-52. For a discussion of how hate speech bans have fared under the ICCPR Article 19(3) tests of legality, legitimacy, and necessity, see Part I.C.

⁶⁹ *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 10.

⁷⁰ *Id.* A heckler’s veto refers to a situation in which a government restricts speech because of the potential or actual reactions of those who disagree with the speech. *See* *Brown v. Louisiana*, 383 U.S. 131, 133 n.1 (1966). In *Brown*, the U.S. Supreme Court highlighted that it had found on three occasions that peaceful African American protesters had been wrongfully charged under Louisiana’s breach of the peace statute because of the potential reaction of others and noted how a scholar had coined the phrase “heckler’s veto” to describe such scenarios. *Id.*

interpreting ICCPR Article 20(2),⁷¹ which culminated in the Rabat Plan of Action.⁷² The Rabat recommendations highlighted the importance of using non-censorial measures as set forth in Resolution 16/18 to deal with intolerance.⁷³ Recognizing that criminalization should be considered only in the most serious cases, the Rabat Plan of Action identified six factors that states should consider when assessing the lawfulness of criminal sanctions: (1) the social and political context when the speech was made; (2) the status of the speaker; (3) the intent of the speaker (noting negligence and recklessness would not suffice); (4) the content and form of the speech; (5) the reach of the speech; and (6) the likelihood of harm, including its imminence.⁷⁴

2. ICERD Article 4

ICERD rests on a premise that racist speech contributes “to creating a climate of racial hatred and discrimination.”⁷⁵ It thus bans the dissemination of ideas based on racial superiority and incitement to racial discrimination and violence. Specifically, with “due regard” for other human rights, it requires state parties to:

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.⁷⁶

Article 4(a) specifies criminal sanctions for speech that falls within its terms.⁷⁷

In 2013, the Committee on the Elimination of Racial Discrimination issued General Comment 35, interpreting Article 4(a) in a way that significantly narrowed the breadth one might have taken from its terms

⁷¹ U.N. High Comm’r for Human Rights, *Rep. on the Expert Workshop on the Prohibition of Incitement to National, Racial, or Religious Hatred*, A/HRC/22/17/Add.4, ¶ 1 (Oct. 5, 2012) (*Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility, or Violence*).

⁷² *Id.* app. ¶¶ 1, 6.

⁷³ *Id.* ¶¶ 23-29, n.4.

⁷⁴ *Id.* ¶ 29. The 2019 Special Rapporteur report to the U.N. General Assembly highlighted the importance of using the Rabat factors in assessing the gravity of various forms of hate speech. See *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 24.

⁷⁵ Committee on the Elimination of Racial Discrimination, General Recommendation No. 35, *Combating Racist Hate Speech*, U.N. Doc. CERD/C/GC/35 ¶ 5 (Sept. 26, 2013) [hereinafter CERD, General Recommendation No. 35].

⁷⁶ ICERD, *supra* note 11, art. 4(a) (emphasis added).

⁷⁷ *Id.* (requiring State Parties to make certain speech “an offence punishable by law,” which implies criminalization rather than solely prohibition.)

only. The Committee took the position that any criminalization of the “dissemination” of racist ideas or “incitement” to harms should include three basic elements: intent, incitement, and particular harms,⁷⁸ similar to the factors developed by the U.N. Special Rapporteur and the Human Rights Committee for the interpretation of ICCPR Article 20.⁷⁹ First, ICERD found that any Article 4 offense should include an examination of the speaker’s specific intent, which involves assessing if the speaker “seeks to influence others to engage in certain forms of conduct . . . through advocacy or threats.”⁸⁰ Former ICERD Committee Member (2001-2014), and lead drafter of General Recommendation 35, Patrick Thornberry has noted that this interpretation “decisively rejects any suggestion of a ‘strict liability’ approach to dissemination and incitement . . . [by linking] them with principles of criminal law on mental elements in crime.”⁸¹ Second, the Committee noted, any speech that is criminalized under Article 4 should rise to the level of incitement, which it defines as “the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech.”⁸² Thornberry has observed that the inclusion of the concept of imminence narrows “the scope of potential hate speech prosecutions.”⁸³ Third, with respect to the harms at stake, the text of Article 4(a) itself lists the harms of discrimination and violence with respect to incitement but sets forth no harms for “dissemination” crimes. In General Recommendation 35, published in 2013, the Committee expanded the list of harms to be averted to include incitement to “hatred” or “contempt,”

⁷⁸ CERD General Recommendation No. 35, *supra* note 75, ¶ 16 (noting that the elements of incitement should apply to the other offenses punishable pursuant to ICERD Article 4, which includes the dissemination of ideas based racial superiority or hatred).

⁷⁹ See *Special Rapporteur 2012 Report to UNGA*, *supra* note 60 and accompanying text. Though the ICERD Committee did not cite the Special Rapporteur’s 2012 report in reaching this conclusion, it did note at the beginning of its General Recommendation that “The Committee has integrated [the] right to freedom of expression into its work on combating hate speech, commenting where appropriate on its lack of effective implementation and, where necessary, drawing upon its elaboration in sister human rights bodies.” CERD General Recommendation No. 35, *supra* note 75, ¶ 4.

⁸⁰ CERD General Recommendation No. 35, *supra* note 75, ¶ 16.

⁸¹ Patrick Thornberry, *International Convention on the Elimination of All Forms of Racial Discrimination: The Prohibition of “Racist Hate Speech,”* in *THE UNITED NATIONS AND FREEDOM OF EXPRESSION AND INFORMATION: CRITICAL PERSPECTIVES* 121, 131 (Cambridge University Press 2015).

⁸² CERD General Recommendation No. 35, *supra* note 75, ¶ 16. The Committee further stated that five contextual factors should be examined before criminalization can be justified for Article 4 offenses: (1) the content and form of the speech; (2) the economic, social, and political climate; (3) the position or status of the speaker; (4) the reach of the speech; and (5) the objectives of the speech, which it adapted from the Rabat Plan of Action. *Id.* ¶ 15, n.17; CERD General Recommendation No. 35, *Combating Racist Hate Speech*, U.N. Doc. No CERD/C/GC/35/Corr.1 (Feb. 14, 2014) (correcting General Recommendation No. 35 to say “[f]or paragraph 14 above read paragraph 15 above”).

⁸³ Thornberry, *supra* note 81, at 132.

which the Committee did not define.⁸⁴ However, Thornberry has noted that such an expansion of harms was tempered by the high thresholds imposed in other parts of General Recommendation 35, including, with respect to intent, incitement to imminent harm, and the application of ICCPR Article 19(3) to any speech restrictions imposed under ICERD Article 4(a).⁸⁵

In addition, General Recommendation 35 “clarified the ‘due regard’ language of ICERD Article 4 to require strict compliance with freedom of expression guarantees.”⁸⁶ General Recommendation 35 states that the due regard clause means the protection of freedom of expression is the “most pertinent principle when calibrating the legitimacy of speech restrictions.”⁸⁷ The ICERD Committee further emphasized that any criminalization of speech under Article 4 “should be governed by the principles of legality, proportionality and necessity.”⁸⁸ Thus, applying the ICCPR Article 19(3) tripartite tests further constrains the potential scope of ICERD Article 4. Indeed, the Committee emphasized that criminalization is only justified in the most serious cases and that expressions of views in academic debates or as political engagement without incitement to harm even when controversial should not be limited.⁸⁹

General Recommendation 35 thus converges with ICCPR Articles 19(3) and 20(2) as interpreted by the Human Rights Committee.⁹⁰ Since issuing General Recommendation 35, the ICERD Committee has not had occasion to adopt substantive views on Article 4(a) through its individual complaints process. Given the Committee’s prior approach to racist expression, however, General Recommendation 35 is likely to significantly impact the Committee’s treatment of hate speech. For example, Thornberry notes that the Committee’s prior recommendations that discounted intent

⁸⁴ CERD General Recommendation No. 35, *supra* note 75, ¶ 13 (b), (d).

⁸⁵ Thornberry, *supra* note 81, at 290-91 (“In practice, any implicit widening of the scope of the offenses to be declared is likely to be countered by the more stringent requirements for the operation of law set out particularly in paragraphs 12, 14, 15, 16, 18, 20 and 25” of General Recommendation 35.) *See infra* note 88 and accompanying text (noting the ICERD Committee’s view that ICCPR Article 19’s tripartite test applies to speech restrictions imposed under ICERD).

⁸⁶ *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 15.

⁸⁷ CERD General Recommendation No. 35, *supra* note 75, ¶ 19.

⁸⁸ *Id.* ¶ 12 (citing to the Human Rights Committee’s General Comment 34).

⁸⁹ *Id.* ¶¶ 12, 25.

⁹⁰ PATRICK THORNBERRY, *THE INTERNATIONAL CONVENTION ON THE ELIMINATION OF ALL FORMS OF RACIAL DISCRIMINATION: A COMMENTARY* 297-98, 301-02 (2016) (“[T]he fresh reading of Article 4 takes the Convention closer to the ICCPR. Overall, it may be argued that General Recommendation 35 takes ICERD practice nearer to ‘libertarian’ currents regarding the prosecution of hate speech crimes; the suggested criminal law requirement of the need for ‘imminence’ of the consequences of incitement may be more stringent than in many jurisdictions”); *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 15 (noting the ICERD Committee’s “converging interpretations” with the Human Rights Committee).

and incitement in hate speech cases are no longer “in accord with current practice as represented by General Recommendation 35.”⁹¹ In addition, before General Recommendation 35, the Committee found violations of ICERD Article 4 by State Parties without assessing the applicability of the ICCPR Article 19(3)’s legality or necessity principles,⁹² which is also no longer in accord with the Committee’s latest interpretations as set forth in General Recommendation 35.

In sum, General Recommendation 35 significantly shifts the ICERD Committee’s approach to hate speech in its most authoritative interpretation of Article 4 and far-reaching integration of ICCPR Article 19. Since adopting the General Recommendation, the ICERD Committee requires the following elements for the criminalization of any speech (including with respect to dissemination of racial superiority and hatred) under Article 4(a): (1) the specific intent of the speaker to influence others to engage in certain conduct; (2) incitement (which is defined as including the imminent risk or likelihood that the intended conduct will result); and (3) the harms of discrimination, violence, hatred, or contempt. In addition to holding “dissemination” to incitement standards and interpreting incitement standards consistent with the high threshold of imminent harm, the Committee also assesses any speech bans imposed under ICERD Article 4 for compliance with ICCPR Article 19(3)’s rigorous tripartite test.

C. *Application of Tripartite Test to Hate Speech Bans*

States may sometimes wish to supplement or substitute an Article 20 argument for banning hate speech with one that rests on assertions that

⁹¹ Thornberry, *supra* note 90, at 294-95. Thornberry refers to the Committee’s 2007 observations on a Ukrainian periodic report in which the Committee was concerned about the “absence of any prosecutions” under its law on “incitement to racial discrimination” and it urged relaxation of the intent standard to “facilitate successful prosecutions.” ICERD Committee, *Concluding Observations of the Committee on the Elimination of Racial Discrimination, Ukraine*, U.N. Doc. CERD/C/UKR/CO/18, ¶ 9 (Feb. 8, 2007). Similarly, in its periodic reporting process prior to General Recommendation 35, the Committee was pleased that Cyprus had removed an intent requirement from its criminal ban on “incitement to racial hatred.” Report of the CERD, GAOR, 56th Sess., Supp. No. 18, A/56/18, ¶ 262 (2001).

⁹² See, e.g., ICERD Committee, *Adan v. Denmark* (Comm’n No. 43/2008), ¶¶ 7.2, 7.7 (Sept. 21, 2010) (finding a violation of Article 4 without assessing whether the relevant law passed legality and necessity conditions—when Denmark “fail[ed] to carry out an effective investigation” into alleged violations of a law that “criminalize[d] public statements by which a group of people are threatened, insulted or degraded on account of their race, colour, national or ethnic origin, religion or sexual inclination”); ICERD Committee, *Jewish Community of Oslo v. Norway* (Comm’n No. 30/2003), ¶¶ 2.5, 10.5 (Aug. 22, 2005) (finding a violation of Article 4—without applying legality and necessity principles—when a conviction had been reversed under a law that restricted “threat[s], insult[s], or subjecting to hatred, persecution or contempt, any person or group of persons because of their creed, race, color or national or ethnic origin”).

particular speech poses a threat to public order or to the rights of others, legitimate grounds for restriction under Article 19(3). Under the ICCPR, whether the basis for restricting speech is the mandatory prohibition under Article 20 or discretionary prohibition under Article 19, states may restrict such hate speech only if they can demonstrate that the restriction meets ICCPR Article 19(3)'s tripartite tests of legality, legitimacy, and necessity.⁹³ U.N. experts have often determined that hate speech laws – whether imposed under obligations to ban hateful speech or as a matter of discretion – fail to meet at least one of these tests.

For instance, laws banning the spread of hateful, divisive, and/or racist views are often inappropriately vague. U.N. monitors have criticized Mauritanian hate speech laws for containing unduly vague prohibitions on speech that (1) has a “racist nature”; (2) “supports or communicates terms that could reveal an intention to leave or incite to hurt morally or physically, promote or incite to hatred”; or (3) “incites discrimination, hatred or violence, defamation and insult on the grounds of origin or belonging racial, ethnic, nationality”⁹⁴ Laws restricting the creation of hatred, discord, or animosity have similarly suffered from vagueness concerns. For instance, in his 2012 report to the U.N. General Assembly, the Special Rapporteur on freedom of expression noted that bans on “causing national, racial, or religious discord and intolerance,” the “expression of feelings of hostility,” and “exciting racial hostility” would be improperly vague.⁹⁵ Malaysian hate speech laws outlawing expression that “conjures feelings of ‘hatred,’ ‘contempt,’ ‘disaffection,’ ‘discontent,’ ‘ill will,’ or ‘hostility’” fail the clarity test.⁹⁶ Moreover, U.N. experts have highlighted that vague hate speech bans “are in fact used to suppress critical and opposing voices” and in Somalia, for example, were used to detain journalists.⁹⁷ Ethiopia’s 2020 criminalization of expression “that deliberately promotes hatred, discrimination or attack against a person or a discernable group of identity, based on ethnicity, religion, race, gender or

⁹³ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶¶ 50-52.

⁹⁴ David Kaye et al. (Special Rapporteur on the Right to Freedom of Opinion and Expression), Mandates of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and of Expression, Comm. No. MRT 5/2017, 8 (Jan. 24, 2018), www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-MRT-5-2017.pdf.

⁹⁵ *Special Rapporteur 2012 Report to UNGA*, *supra* note 60, ¶ 51.

⁹⁶ David Kaye (Special Rapporteur on the Right to Freedom of Opinion and Expression), *Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, Comm. No. OL MYS 6/2018, 3 (Dec. 28, 2018), <https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gId=24287>.

⁹⁷ *Special Rapporteur 2012 Report to UNGA*, *supra* note 60, ¶ 51 (highlighting misuse of hate speech laws in Somalia “to arrest and detain independent journalists”).

disability” involved “a serious risk that the law may be used to silence critics.”⁹⁸

As part of the U.N.’s pivot towards increasingly protective approaches to freedom of expression, the Human Rights Committee emphasized in General Comment 34 that certain governmental objectives do not meet ICCPR Article 19’s “legitimacy” test. For example, the Committee stated that prohibiting expression that denies historic facts is not a legitimate reason to burden speech.⁹⁹ As noted by General Comment 34’s lead drafter, this position represented a purposeful reversal of the approach the Committee had taken in a prior matter.¹⁰⁰ The Committee did note that any limitations on speech on the denial of historic facts should comply with ICCPR Articles 19 and 20, which indicates that governments may prohibit expression that denies historic facts, but only when such speech constitutes advocacy of incitement to particular harms. Similarly, in General Recommendation 35, the ICERD Committee emphasized that “the expression of opinions about historical facts” should not be banned,¹⁰¹ and should be criminalized only where it rises to the level of incitement.¹⁰²

The Human Rights Committee has also made clear that laws that have the objective of mandating respect for religion or religious leaders are not acceptable governmental goals.¹⁰³ As noted by one of the authors, “anti-blasphemy laws fail to meet the legitimacy condition of Article 19(3), given that Article 19 protects *individuals* and their rights to freedom of expression and opinion; neither it nor Article 18 of the ICCPR protect ideas or beliefs from ridicule, abuse, criticism or other ‘attacks’ seen as offensive.”¹⁰⁴ That said, if the blasphemous speech is part of advocacy of religious hatred that rises to the level of incitement to violence, discrimination, or hostility against that group, then it would be subject to

⁹⁸ David Kaye (Special Rapporteur on 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: Visit to Ethiopia*, U.N. Doc. No A/HRC/44/49/Add.1, ¶¶ 33-34 (Apr. 29, 2020).

⁹⁹ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶ 49:

Laws that penalize the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression. The Covenant does not permit general prohibition of expressions of an erroneous opinion or an incorrect interpretation of past events.

¹⁰⁰ O’Flaherty, *supra* note 33, at 653 (referring to HRC, *Faurisson v. France* (Communication no. 550/1993), 16 December 1996, § 9.5). In *Faurisson*, France had convicted a man under the Gayssot Act for disagreeing with the conclusions of the International Military Tribunal at Nuremberg. The Committee had upheld that conviction under ICCPR Article 19. *Id.*

¹⁰¹ CERD General Recommendation No. 35, *supra* note 75, ¶ 14 (citing to U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶ 49).

¹⁰² *Id.*

¹⁰³ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶ 48.

¹⁰⁴ *Special Rapporteur 2019 Report to UNGA*, *supra* note 14, ¶ 21.

prohibition under ICCPR Article 20 (assuming compliance with Article 19(3)'s three-part test).¹⁰⁵

With respect to the necessity test, U.N. mechanisms have advocated that non-censorial methods to combat hate speech should be considered, consistent with the approach embodied in Resolution 16/18.¹⁰⁶ Where such measures are insufficient, a government should assess its range of options that burden speech and select the least intrusive means to achieve its legitimate purpose (such as the rights of others or public order).¹⁰⁷ International human rights law underscores the importance of demonstrating a likely near-term (i.e., imminent) harm to justify a speech restriction as the least intrusive means for achieving a legitimate objective.¹⁰⁸ U.N. monitors often admonish states for neglecting the incitement standard and its various contextual factors, including the likelihood of imminent harm, when imposing hate speech bans.¹⁰⁹ In addition, the global human rights framework's requirement that the intention of the speaker to cause particular harm be examined constitutes a further and substantial threshold in restricting hate speech.¹¹⁰ Given the Special Rapporteur's understandings of the thresholds implicit in incitement that were described above,¹¹¹ the least intrusive means test compels the state to approach the imposition of hate speech bans in a disciplined and calibrated manner.

¹⁰⁵ *Id.*

¹⁰⁶ *Id.* ¶¶ 18-24 (“The recognition of steps other than legal prohibitions highlights that prohibition will often not be the least restrictive measure available to States confronting hate speech problems.”); *Special Rapporteur 2012 Report to UNGA*, *supra* note 60, ¶¶ 62-64.

¹⁰⁷ *See supra* notes 46-51 (describing ICCPR Article 19(3)'s least intrusive means test).

¹⁰⁸ Frank LaRue (Special Rapporteur on 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*, U.N. Doc. No A/68/362, §§ 52, 53 (Sept. 4, 2013):

For a restriction to be necessary, it must . . . not be more restrictive than is required for the achievement of the desired purpose or protected right [T]he authorities must demonstrate, in specific and individualized fashion, the precise nature of the *imminent* threat, as well as the necessity for and the proportionality of the specific action taken. A direct and immediate connection between the expression (or the information to be disclosed) and the alleged threat must be established.

¹⁰⁹ *Special Rapporteur 2012 Report to UNGA*, *supra* note 60, ¶ 79.

¹¹⁰ *Id.* ¶ 50 (citing a 2001 joint statement with the freedom of expression experts from the Organization of American States and the Organization for Security and Cooperation in Europe) (“No one should be penalized for the dissemination of hate speech unless it has been shown that they did so with the intention of inciting discrimination, hostility or violence”).

¹¹¹ *See supra* notes 60-65 and accompanying text (requiring speakers display a specific intent to cause harm, defining incitement as creating an imminent risk of harm, and viewing hostility as “a manifestation of hatred beyond a mere state of mind”).

II. REGIONAL HUMAN RIGHTS STANDARDS ON EXPRESSION AND HATE SPEECH

Although the U.N.'s global human rights framework has converged around an interpretation of those provisions protecting and promoting freedom of expression while also requiring prohibitions of 'hate speech,' regional jurisprudence has not always arrived at the same conclusions. Part II examines areas of convergence and conflict, paying particular attention to the central regional institutions that have developed legal instruments and jurisprudence around the freedom of expression. The most prolific regional jurisprudence on hate speech derives from the European Court of Human Rights. The courts within the American and African systems do not have substantial jurisprudence on hate speech and thus our analysis of those regional approaches is often based on interpretations by authoritative expert (but not court) entities within those systems. As the European, American, and African approaches are the most developed regional human rights systems, we devote the most attention to the central human rights treaties in those regions.

In particular, Part II highlights how the European Court of Human Rights has alleviated the burden on governments to justify speech restrictions in a variety of ways that provide fewer protections for freedom of expression than the U.N. system.¹¹² We note how this body of European case law is frequently relied upon by other regional mechanisms, which risks leading other regions to depart from the U.N.'s minimum human rights protections.¹¹³ In addition, this Part identifies unfolding conflicts with the U.N.'s standards in the emergent, weakly-institutionalized approaches of the Middle East, the Association of Southeast Asian Nations (ASEAN), and the OIC.¹¹⁴ Again, we find the European Court's jurisprudence invoked by one of these evolving approaches (the OIC) to justify providing fewer protections for freedom of expression than what is afforded under global standards.¹¹⁵

A. *The Inter-American Human Rights System*

The American Convention on Human Rights was opened for signature in 1969, entered into force in 1978, and currently has twenty-five state

¹¹² See *infra* notes 138-90 and accompanying text for a discussion of the European system.

¹¹³ See, e.g., *infra* notes 198-204 and accompanying text (describing how the African Court has adopted certain approaches from the European Court that depart from U.N. standards).

¹¹⁴ See *infra* Part II(D) and accompanying text.

¹¹⁵ See *infra* note 232 and accompanying text.

parties.¹¹⁶ The American Convention provides for broad protections for freedom of expression, including an explicit prohibition on prior censorship.¹¹⁷ In terms very similar to ICCPR Article 19, American Convention Article 13 protects the “freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other medium of one’s choice.”¹¹⁸ Governments may not restrict freedom of expression via prior censorship, but may restrict it through “subsequent imposition of liability” when (1) “expressly established by law” and (2) necessary (3) “to ensure” “(a) respect for the rights or reputations of others; or (b) the protection of national security, public order, or public health or morals.”¹¹⁹ The Inter-American system has a presumption in favor of speech and places the burden on the government to demonstrate that any hate speech ban satisfies its strict conditions of legality, necessity, and legitimacy when restricting hate speech.¹²⁰

The Inter-American Special Rapporteur on Freedom of Expression (IACHR Special Rapporteur), whose findings carry significant weight within the Inter-American Commission on Human Rights, is based on the American Convention and has synthesized relevant interpretations of the three-part tests of legality, legitimacy, and necessity in a manner that converges in material respects with the U.N.’s approach. Specifically, limitations on speech must be defined in a precise and clear manner by a law,¹²¹ and the limitation must serve “compelling” objectives set forth in the

¹¹⁶ Organization of American States, American Convention on Human Rights, Nov. 22, 1969, O.A.S.T.S. No. 36, 1144 U.N.T.S. 123 [hereinafter American Convention]; *Signatories and Current Status of Ratifications*, INTER-AM. COMM’N H.R., <https://www.cidh.oas.org/basicos/english/Basic4.Amer.Conv.Ratif.htm> (last visited Jan. 12, 2022).

¹¹⁷ American Convention, *supra* note 116, arts. 13(1)-(2). Prior censorship refers to governmental suppression of speech before it can be disseminated to the public. *Prior Censorship*, THE ORG. OF AM. STATES, <https://www.oas.org/en/iachr/expression/showarticle.asp?artID=152&IID=1> (describing Inter-American human rights bodies’ interpretation of Article 13’s prohibition of prior censorship).

¹¹⁸ American Convention, *supra* note 116, art. 13(1).

¹¹⁹ *Id.* art. 13(2). It should be noted that a subsequent provision does temper the prohibition on prior censorship in limited circumstances. *Id.* art. 13(4) (“public entertainments may be subject by law to prior censorship for the sole purpose of regulating access to them for the moral protection of childhood and adolescence.”).

¹²⁰ Inter-Am. Comm’n H.R., *The Inter-American Legal Framework Regarding the Right to Freedom of Expression*, ¶ 68, OEA/Ser.L/V/II CIDH/RELE/INF. 2/09 (Dec. 30, 2009), <http://www.oas.org/en/iachr/expression/docs/publications/INTER-AMERICAN%20LEGAL%20FRAMEWORK%20OF%20THE%20RIGHT%20TO%20FREEDOM%20OF%20EXPRESSION%20FINAL%20PORTADA.pdf> (“It is incumbent upon the authority imposing the limitations to prove that these conditions have been met. Furthermore, all of the stated conditions must be met simultaneously in order for the limitations to be legitimate pursuant to the American Convention”).

¹²¹ *Id.* ¶¶ 69-70:

American Convention.¹²² Indeed, “necessary” is not synonymous with “useful,” “reasonable,” or “convenient.”¹²³ Rather, to meet the necessity test, states must demonstrate they have selected the least restrictive means of achieving the public interest objective.¹²⁴ In addition, states must show the speech restriction is “strictly proportionate” to the objective pursued, which involves an assessment of “whether the sacrifice of freedom of expression . . . is excessive in relation to the advantages obtained through such measure.”¹²⁵ With respect to any speech restriction, the burden is on “the authority imposing limitations to prove that these conditions have been met.”¹²⁶

Of the regional human rights systems, only the Inter-American system has a mandatory requirement to criminalize certain hate speech in its foundational human rights instrument. American Convention Article 13(5), with language similar to ICCPR Article 20, requires that “any *advocacy of national, racial, or religious hatred that constitute[s] incitements to lawless violence or to any other similar action against any person or group of persons on any grounds including those of race, color, religion, language, or national origin shall be considered as offenses punishable by law.*”¹²⁷ For speech to rise to this level, the expression therefore must contain several elements: (1) advocacy of national, racial, or religious hatred that rises to the level of (2) incitement to (3) lawless violence or similar harm (4) against any person or group.

The Inter-American Court of Human Rights (IACHR) has not yet opined on the scope of Article 13(5). However, the IACHR Special Rapporteur has assessed its meaning. With respect to the intent element, the

Every limitation to freedom of expression must be established in advance, expressly, restrictively and clearly in a law . . . The laws that set limits to freedom of expression must be drafted in the clearest and most specific terms possible, as the legal framework must provide legal certainty to the public.

The Inter-American Special Rapporteur has noted the chilling effect of vague speech bans as well as the inappropriate discretion they grant to governmental enforcers. *Id.* ¶ 71 (“Vague, ambiguous, broad or open-ended laws, by their mere existence, discourage the dissemination of information and opinions out of fear of punishment, and can lead to broad judicial interpretations that unduly restrict freedom of expression”).

¹²² *Id.* ¶ 74 (“[R]estrictions imposed must pursue one of the limited compelling objectives set forth in the American Convention”).

¹²³ *Id.* ¶ 85.

¹²⁴ *Id.* ¶ 86 (“[A]mong the various options available for reaching the same objective, the State should choose the one that least restricts [freedom of expression]”). The state must also show that the restriction is a measure that is “effectively conducive to attaining the legitimate and compelling objectives in question.” *Id.* ¶ 87.

¹²⁵ *Id.* ¶ 88.

¹²⁶ *Id.* ¶ 68.

¹²⁷ American Convention, *supra* note 116, art. 13(5) (emphasis added).

IACHR Special Rapporteur found that there must be advocacy of hatred against a particular group, requiring evidence that the speaker had the intention and capacity to engage in an unlawful act.¹²⁸ In other words, the speaker must not just intend to speak, but instead must specifically intend for harm to occur and must have the capacity to create or incite harm. The Inter-American Special Rapporteur has also observed that, because the phrasing of the treaty provision covers harm to groups beyond those based on nationality, race, and religion, the targeting of unlisted groups such as “individuals on the basis of their sexual orientation, gender identity, or bodily diversity” would also be covered by this provision.¹²⁹

In addition, the speech must rise to the level of “incitement.” The Inter-American Special Rapporteur explained that this means it is impermissible to limit freedom of expression based on “mere conjectures about eventual effects on order, nor hypothetical circumstances derived from subjective interpretations by authorities of facts.”¹³⁰ Instead, the state must demonstrate facts that “clearly present a present, certain, objective, and *imminent* risk of violence.”¹³¹ With regard to the types of harms that are targeted in this provision, the advocacy of incitement must be directed to producing “lawless violence” or “similar [illegal] action.”¹³² Though the scope of “similar action” is not defined, the phrasing indicates the harm should have the same severity as violence. In comparison to the U.N. system, the American Convention shares similar intent and incitement standards, but the types of harm appear broader in ICCPR Article 20.¹³³

When hateful speech does not rise to the level of American Convention Article 13(5), the Inter-American system provides that such speech *may* be restricted, but only if (1) its three-part test of legality, legitimacy, and necessity is satisfied, and (2) the burden on speech takes

¹²⁸ Catalina Botero Marino (Inter-Am. Comm’n H.R. Special Rapporteur for freedom of expression), Annual Report of the Inter-Am. Comm’n H.R., OEA/Ser.L/V/II. Doc. 51, ¶ 544 (Dec. 20, 2009), <http://www.oas.org/en/iachr/expression/docs/reports/annual/Informe%20Anual%202009%202%20ENG.pdf> [hereinafter Botero 2009 Report] (“these must have as a prerequisite strong, objective evidence that the person was not simply expressing an opinion, but also had the clear intention to commit an unlawful act and the real, present, and effective possibility of achieving his or her objectives.”).

¹²⁹ Inter-Am. Comm’n H.R., *Hate Speech and Incitement to Violence Against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas*, ch. IV, ¶ 17 (Nov. 12, 2015), http://www.oas.org/en/iachr/expression/docs/reports/hate/Hate_Speech_Incitement_Violence_Against_LGBTI.pdf [hereinafter *2015 IACHR Report*].

¹³⁰ Botero 2009 Report, *supra* note 128, ¶ 524.

¹³¹ *Id.* ¶ 544 (emphasis added).

¹³² American Convention, *supra* note 116, art. 13(5).

¹³³ See *supra* notes 59-65 and accompanying text (describing the harms encompassed under ICCPR Article 20 as discrimination, hostility, and violence, but noting the Special Rapporteur’s restrictive understanding “hostility”).

the form of subsequent sanctions rather than prior restraints.¹³⁴ The Inter-American Special Rapporteur has also cautioned that legal prohibitions may not be effective to counter hateful speech and that, instead, states should deploy a more holistic approach that involves a variety of good governance measures. This echoes calls at the U.N. level to seek use of non-censorial methods first to combat intolerance.¹³⁵ Specifically:

[n]egative or derogatory portrayal and other expressions that stigmatize LGBTI persons are certainly offensive and hurtful and they increase the marginalization, stigmatization, and general insecurity of LGBTI persons. However, the IACHR is of the opinion that the legal prohibition of this type of speech will not do away with the stigma, prejudice, and hatred against LGBTI persons that is deeply rooted in the societies of the Americas. In many contexts, given the structural social inequalities, discriminatory views and prejudice in public discourse cannot be eradicated through legal sanctions [M]ore should be done to promote a comprehensive approach that goes beyond legal measures and includes preventive and educational mechanisms and measures implemented by States, media, and society in general.¹³⁶

As noted above, the Inter-American approach also imposes strict tests of legality, necessity, and legitimacy, including by requiring governments to prove that they selected the “least intrusive means.”¹³⁷

B. The European Human Rights System

The European Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention), which was adopted in

¹³⁴ 2015 *IACHR Report*, *supra* note 129, ¶ 20 (“as with any other restriction on freedom of expression, the imposition of subsequent sanctions must satisfy the requirements set forth in Article 13(2) of the American Convention and be applied by an independent state entity.”). It should be noted that in 2020 the InterAmerican Convention Against All Forms of Discrimination and Intolerance entered into force when two countries submitted their instruments of ratification. *Signatories and Ratifications*, INTER-AM. CONVENTION AGAINST ALL FORMS OF DISCRIMINATION & INTOLERANCE, https://www.oas.org/en/sla/dil/inter_american_treaties_a-69_discrimination_intolerance_signatories.asp (last visited Jan. 12, 2022); Organization of American States, Convention Against All Forms of Discrimination and Intolerance, O.A.S.T.S. No. __, ___U.N.T.S. __. Although this treaty contains facially broad bans on intolerant speech in Article 4, one would presume that the Inter-American system would continue to apply its rigorous and principled approach to legality, legitimacy, and necessity as well as its high thresholds for incitement when interpreting these provisions.

¹³⁵ See *supra* notes 47-52, 106 and accompanying text (describing the U.N.’s endorsement of a toolkit of governmental measures that should be deployed to tackle hate and intolerance before resorting to speech bans).

¹³⁶ 2015 *IACHR Report*, *supra* note 129, ¶ 21.

¹³⁷ See Inter-Am Comm’n H.R., *supra* note 120, ¶ 84-85 and accompanying text.

1950, entered into force in 1953, and has 47 state parties,¹³⁸ provides the foundation for the most developed human rights adjudicatory system worldwide. Though the European Convention's protection of freedom of expression echoed UDHR Article 19 and preceded ICCPR Article 19, the European Court of Human Rights (European Court) has developed a jurisprudential approach which varies in a number of respects from the U.N.'s approach, with particular salience in the context of hate speech.¹³⁹ Article 10 of the European Convention provides that freedom of expression "shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers."¹⁴⁰ Article 10(2), similar to ICCPR Article 19(3), permits restrictions when a three-part test is met: the restrictions must be (1) prescribed by law and (2) necessary in a democratic society for (3) national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.¹⁴¹

Despite similar phrasing to the ICCPR, the European Court has diminished the burden on governments to demonstrate the validity of hate speech bans. It has, for example, taken the position that certain hate speech falls outside the scope of Article 10's protections if the speaker is using freedom of expression to undermine other rights in the Convention.¹⁴²

¹³⁸ Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, *opened for signature* Nov. 4, 1950, 213 U.N.T.S. 221 (entered into force Sept. 3, 1953) [hereinafter ECHR].

¹³⁹ This Article focuses on an examination of the over forty cases that are highlighted in the European Court of Human Rights, September 2020 Fact Sheet on Hate Speech. EUR. CT. H.R., FACT SHEET – HATE SPEECH (Sep. 2020), <https://perma.cc/37QL-KXM7> [hereinafter ECtHR Fact Sheet on Hate Speech].

¹⁴⁰ ECHR, *supra* note 138, art. 10. This article also provides that it "shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises." *Id.*

¹⁴¹ *Id.*

¹⁴² See *Lilliendahl v. Iceland*, App. No. 29297/18, ¶¶ 33-40 (May 12, 2020), <http://hudoc.echr.coe.int/eng?i=001-203199>. In this case, the European Court explained that its jurisprudence on hate speech falls into two categories. *Id.* ¶ 33. The first category "is comprised of the gravest forms of 'hate speech,' which the Court has considered fall under Article 17 and thus [are] excluded entirely from the protection of Article 10." *Id.* ¶ 34. ECHR Article 17 provides that

[n]othing 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.

ECHR, *supra* note 138, art. 17. When the European Court determines that speech falls in this category, it finds the appeal inadmissible, thereby relieving the government of demonstrating the validity of the hate speech ban under Article 10, which effectively leaves the national penalty intact. See, e.g.,

When confronted with such speech, the European Court finds the case inadmissible without reaching the merits, leaving in place the challenged national penalty for such expression.¹⁴³ Examples of hate speech that reach this level include: the possession and intention to distribute materials that called on immigrants to depart the Netherlands;¹⁴⁴ display of a poster from a home window that showed the Twin Towers in flames and stated “Islam out of Britain—Protect British People;”¹⁴⁵ upload of YouTube videos that called for Muslims to fight non-Muslims;¹⁴⁶ and publication of anti-Semitic statements by a newspaper owner and editor.¹⁴⁷ In addition, the European Court has determined that a variety of forms of Holocaust denial do not fall within the protection of the European Convention.¹⁴⁸

When it identifies “less grave” forms of hate speech, the European Court examines whether the government’s intrusion on speech is consistent with the European Convention’s protection for freedom of expression.¹⁴⁹ However, the European Court frequently relies on a doctrine known as the “margin of appreciation,”¹⁵⁰ under which it defers to governmental judgments, including when states disagree about the scope of a right.¹⁵¹ The European Court affords governments a particularly wide margin of appreciation when expression offends religious or moral convictions, when competing rights need to be balanced, or when there is the potential for

Norwood v. the United Kingdom, 2004-XI Eur. Ct. H.R. 342, 343 (2004), <http://hudoc.echr.coe.int/eng?i=001-67632> (finding applicant’s appeal inadmissible because the speech in question contravened Article 17). The second category of hate speech “is comprised of ‘less grave’ forms of ‘hate speech’ which the Court has not considered to fall entirely outside the protection of Article 10, but which it has considered permissible for Contracting States to restrict.” Lilliendahl, App. No. 29292/18, ¶ 35 (May 12, 2020). For hate speech that falls into the second category, the Court will assess if Article 10(2)’s three-part test is met. *Id.* ¶¶ 39-48 (analyzing whether the legality, legitimacy, and necessity tests were met for hate speech that fell within the second category).

¹⁴³ For examples of particular cases that meet this threshold, see *infra* notes 144-48 and accompanying text.

¹⁴⁴ Glimmerveen v. The Netherlands, App. Nos. 8348/78 & 8406/78, 18 Eur. Comm’n H.R. Dec. & Rep. 187, 188-97 (1979).

¹⁴⁵ Norwood, 2004-XI Eur. Ct. H.R. at 343.

¹⁴⁶ Belkacem v. Belgium, App. No. 34367/14, ¶¶ 5-7 (June 27, 2016), <http://hudoc.echr.coe.int/eng?i=001-175941>.

¹⁴⁷ Ivanov v. Russia, App. No. 35222/04, 1 (Feb. 20, 2007), <http://hudoc.echr.coe.int/eng?i=001-79619>.

¹⁴⁸ *Id.*; M’Bala v. France, App. No. 25239/13, ¶ 39 (Oct. 20, 2015), <http://hudoc.echr.coe.int/eng?i=001-158752>; Garaudy v. France, 2003-IX Eur. Ct. H.R. 369, 1, 23.

¹⁴⁹ Lilliendahl, App. No. 29297/18, ¶ 35-36.

¹⁵⁰ PHILIP ALSTON & RYAN GOODMAN, INTERNATIONAL HUMAN RIGHTS 946-48 (2012) (explaining the “margin of appreciation”).

¹⁵¹ See *id.* at 947.

violence.¹⁵² However, the European Court will narrow the margin of appreciation in situations where the speaker is discussing certain matters of public concern.¹⁵³ In 2011, the U.N. Human Rights Committee unequivocally rejected any application of the margin of appreciation with regard to freedom of expression.¹⁵⁴

The European Court has also invoked the margin of appreciation in deferring to governmental assessments about the appropriateness of hate speech convictions. For example, in the *Soulas* case, which involved France imposing criminal penalties on those responsible for the publication of an anti-immigrant book,¹⁵⁵ the European Court noted that speech restrictions must be convincingly established, but observed that dealing with issues of immigration and integration “is up to the national authorities, who have a deep understanding of the realities of the country. They must therefore have a sufficiently wide margin of appreciation to determine . . . the need for such interference [with expression].”¹⁵⁶ As another example, in sustaining Russia’s conviction of a journalist for inciting hatred, the Court relied on its margin of appreciation doctrine to accept that the speaker’s article could stir up “base emotions or embedded prejudices” against those

¹⁵² See, e.g., *I.A. v. Turkey*, 2005-VIII Eur. Ct. H.R. 235, ¶ 25 (2005); *Tagiyev v. Azerbaijan*, App. No. 13274/08, ¶ 39 (Dec. 5, 2019), <http://hudoc.echr.coe.int/eng?i=001-198705>; *Delfi AS v. Estonia*, App. No. 64569/09, ¶ 139 (June 16, 2015), <https://hudoc.echr.coe.int/eng?i=001-155105>; *Magyar Tartalomszolgáltatók Egyesülete v. Hungary*, App. No. 22947/13, ¶ 59 (Feb. 2, 2016), <http://hudoc.echr.coe.int/fre?i=001-160314>; *Terentyev v. Russia*, App. No. 10692/09, ¶ 65 (Aug. 28, 2018), <https://hudoc.echr.coe.int/eng?i=001-185307>.

¹⁵³ See, e.g., *Mondragon v. Spain*, 2011-I Eur. Ct. H.R. 835, ¶ 51 (finding that an elected official’s speech on matters involving his government’s welcome of a head of state and related allegations of ill-treatment constituted an issue of public interest that narrows the margin of appreciation); *Tagiyev*, App. No. 13274/08, ¶¶ 37-39 (Dec. 5, 2019) (noting that a reduced margin of appreciation applies for expression concerning public interest matters, but a wide margin of appreciation is afforded for “matters liable to offend personal convictions within the sphere of morals or religion”).

¹⁵⁴ U.N. Human Rights Committee, General Comment No. 34, *supra* note 30, ¶ 36 (observing that the scope of freedom of expression is not to be assessed using the margin of appreciation doctrine).

¹⁵⁵ *Soulas v. France*, App. No. 15948/03, ¶¶ 15, 43-44, 47-48 (July 10, 2008), <http://hudoc.echr.coe.int/eng?i=001-87370>. In this case, the ECtHR upheld a French conviction for incitement to hatred and violence against a particular group of people. *Id.* ¶ 48. The complainants were those involved with the publication of a book titled “Colonization of Europe.” *Id.* ¶ 6. Under the title appeared the following: “True Discourse on Immigration and Islam.” *Id.* The French court had found the book contained punishable speech because, among other things, it described Muslim immigrants in France

as rebels against the application of the laws in force in France and generally indulging in the practice of criminal activities, not only for the purpose of profit arising from their fundamental inability to integrate into normal economic life but also, being animated by a feeling of anti-European racism, with a view to conquering the territory to exclude the population of European origin and ensure both their power and the hegemonic implantation of the Muslim religion.

Id.

¹⁵⁶ *Id.* ¶¶ 35, 38.

of non-Russian ethnicity, which justified governmental intervention.¹⁵⁷ And again, in upholding Turkey's criminal fines on a managing director of a publishing house for blasphemy, the Court emphasized the particular importance of the margin of appreciation's role.¹⁵⁸

In addition to this deference, the European Court differs significantly from the U.N. and Inter-American systems when applying the legality, legitimacy, and necessity conditions. The European Court has explained that the legality (or "prescribed by law") test means that the interference with speech has a basis in domestic law, is accessible to the person concerned who must be able to foresee its consequences, and is compatible with the rule of law.¹⁵⁹ In a 2016 report, one of the authors highlighted vagueness problems with the European human rights system in particular.¹⁶⁰ A review of leading hate speech cases reveals three trends that are relevant to this concern about the European Court's approach to vagueness. First, the Court frequently has found that hate speech restrictions which would likely fail for vagueness and/or overbreadth under international standards nevertheless meet the European Convention's legality test.¹⁶¹ For instance, the Court held that a law in Bosnia and Herzegovina which criminalized inciting or stirring up "national, racial, or religious hatred, discord, or intolerance" satisfied the legality test.¹⁶² Similarly, the Court held Sweden's penal code, which punishes expressions of contempt against particular groups, met the legality condition.¹⁶³ The Court also determined that Iceland's criminalization of "anyone who publicly mocks, defames, denigrates or threatens a person or group of persons" with certain characteristics passed the legality test.¹⁶⁴ Such restrictions would likely fail international standards due to vagueness or overbreadth, absent clear interpretative guidance on their scope.

Second, the European Court has dismissed vagueness arguments as improperly raised under the legality test, instead considering them as part of the necessity condition. For example, when a member of parliament

¹⁵⁷ *Atamanchuk v. Russia*, App. No. 4493/11, ¶ 64 (Feb. 11, 2020), <http://hudoc.echr.coe.int/eng?i=001-200839>.

¹⁵⁸ *I.A. v. Turkey*, 2005-VIII Eur. Ct. H.R. ¶ 25.

¹⁵⁹ *Dink v. Turkey*, App. No. 2668/07, ¶ 114 (Sept. 14, 2020), <http://hudoc.echr.coe.int/eng?i=001-100383>.

¹⁶⁰ *Special Rapporteur 2016 Report to UNGA*, *supra* note 31, ¶ 25.

¹⁶¹ For a discussion of the U.N. human rights machinery's application of the vagueness test, see *supra* notes 38-39, 94-98 and accompanying text.

¹⁶² *Smajić v. Bosnia and Herzegovina*, App. No. 48657/16, ¶ 13 (Jan. 16, 2018), <http://hudoc.echr.coe.int/eng?i=001-180956>.

¹⁶³ *Vejdeland v. Sweden*, App. No. 1813/07, ¶ 49 (Feb. 9, 2012), <http://hudoc.echr.coe.int/eng?i=01-109046>.

¹⁶⁴ *Lilliendahl*, App. No. 29297/18, ¶¶ 20, 42 (May 12, 2020).

argued that his criminal conviction for “serious insult” against Spain’s king was based on a vague law, the Court held that the law passed the legality test and stated it would address the vagueness concerns in its analysis of whether the interference with speech was proportionate to the state’s aim.¹⁶⁵ In another case, a person convicted under Russia’s penal code for inciting hatred “or enmity and humiliating the dignity of an individual or a group of individuals on the grounds of their membership of a social group” argued that he could not have foreseen the law being applied to the police as a “social group.”¹⁶⁶ The Court again declined to consider this vagueness argument under the legality condition and instead explained it would assess this issue when considering whether the punishment was proportionate to the state’s aim.¹⁶⁷ Under international standards, collapsing vagueness with proportionality is inappropriate, as governments bear the burden of proving that each prong of the legality, legitimacy, and proportionality test is met.¹⁶⁸ When it collapses the tests, the European Court gives short shrift to legality/vagueness as an independent protection against unlawful speech restrictions and weakens its version of the three-part test.

Third, speakers frequently do not even contest the legality condition, including vagueness, in the hate speech cases that they appeal to the European Court. For example, in another case involving a conviction for serious insult to Spain’s king, the speakers conceded that the legality test was met.¹⁶⁹ Similarly, a journalist convicted under a Danish law that criminalizes “threatening, insulting or degrading a group of persons on account of their race, colour, national or ethnic origin or belief” conceded that the legality condition was satisfied.¹⁷⁰ In many other cases, the speakers also did not contest speech restrictions based on vagueness, likely because of the Court’s tolerance for vague wording in hate speech bans.¹⁷¹

With respect to the legitimacy test, the European Court has found that certain public interest objectives are sufficiently important to survive this test despite the fact that they would not survive scrutiny in the U.N. system.

¹⁶⁵ Mondragon, 2011-I Eur. Ct. H.R. 215, ¶¶ 35, 45-46.

¹⁶⁶ Terentyev, App. No. 10692/09, ¶¶ 30, 42 (Aug. 28, 2018).

¹⁶⁷ *Id.* ¶ 58.

¹⁶⁸ See *supra* notes 38-55 and accompanying text.

¹⁶⁹ Taulats, App. No. 51168/15, ¶ 25 (Mar. 13, 2018), <http://hudoc.echr.coe.int/eng?i=001-181719>.

¹⁷⁰ Jersild v. Denmark, App. No. 15890/89, ¶¶ 19, 27 (Sept. 23, 1994), <http://hudoc.echr.coe.int/eng?i=001-57891>.

¹⁷¹ See, e.g., Lehideux v. France, App. No. 55/1997/839/1045 ¶ 39 (Sept. 23, 1998), <http://hudoc.echr.coe.int/eng?i=001-58245>; Altintas v. Turkey, App. No. 50495/08, ¶ 27 (Mar. 10, 2020), <http://hudoc.echr.coe.int/eng?i=001-201897>; I.A. v. Turkey, 2005-VIII Eur. Ct. H.R., ¶ 22; Soulas, App. No. 15948/03, ¶ 26 (July 10, 2008); Erbakan v. Turkey, App. No. 59405/00, ¶ 45 (July 6, 2006), <http://hudoc.echr.coe.int/eng?i=001-76232>.

For example, the European Court has determined that the goal of not offending religious sensibilities is a sufficient governmental interest to justify restricting speech.¹⁷² Though not in the context of a hate speech case, the European Court has also found that the governmental goal of having people live together in society is an appropriate public interest objective to pursue when limiting rights.¹⁷³

As to the necessity test, commentators have observed that the European Court does not apply a “least intrusive means” test, but rather focuses solely on whether the interference with speech is proportionate.¹⁷⁴ A review of the cases listed in the European Court’s 2020 Hate Speech Fact Sheet is consistent with this observation.¹⁷⁵ Overall, the Court did not require the government to demonstrate that the interference with expression was the least intrusive means of achieving its legitimate public interest objective.¹⁷⁶

Instead, the European Court focused on whether the interference with speech was proportionate in these cases. It engaged in its proportionality assessment through a three prong inquiry: whether (1) there is a pressing social need for the interference with speech, (2) the interference is proportionate to the legitimate aim pursued, and (3) the reasons provided by the government are relevant and sufficient.¹⁷⁷ In this inquiry, the Court often notes that it is balancing freedom of expression with other rights or

¹⁷² See, e.g., *Otto-Preminger-Institut v. Austria*, 295 Eur. Ct. H.R. (ser. A) (1994) (determining that the goal of protecting citizens from having their religious feelings insulted constituted a legitimate public aim). In *Otto Preminger*, Austria seized a film that risked offending the religious sensibilities of Christians in an area that was over 85% Roman Catholic. *Id.* ¶ 52. The Court found that “in the context of religious opinions and beliefs” there is an obligation on speakers “to avoid as far as possible expressions that are gratuitously offensive to others and thus an infringement of their rights.” *Id.* ¶ 49. The Court viewed the film’s offensive expression as requiring it to balance freedom of expression interests with “the right of other persons to proper respect for their freedom of thought, conscience and religion.” *Id.* ¶ 55. The U.N. system, on the other hand, has consistently opined that neither freedom of religion nor freedom of expression “protect ideas or beliefs from ridicule, abuse, criticism or other ‘attacks.’” See *supra* note 104 and accompanying text. Similarly, it should be noted that the Inter-American Court of Human Rights overturned Chile’s prior censorship of a film that risked offending religious sensibilities. See *The Last Temptation of Christ v. Chile*, Merits, Reparations, and Costs, Judgment, Inter-Am Ct. H.R. (Ser. C) No. 73 (2001) (Feb. 5, 2001).

¹⁷³ See *Dakir v. Belgium*, App. No. 4619/12, ¶ 57 (July 11, 2017), <https://hudoc.echr.coe.int/eng/?i=001-175660> (finding that the public interest objective of having people live together in society is a legitimate governmental reason to ban burqas).

¹⁷⁴ See, e.g., JONAS CHRISTOFFERSEN, *FAIR BALANCE: PROPORTIONALITY, SUBSIDIARITY AND PRIMACY IN THE EUROPEAN CONVENTION ON HUMAN RIGHTS* 114 (2009) (noting the ECtHR’s “general rejection of the less and least onerous means-test” in its jurisprudence).

¹⁷⁵ See ECtHR Fact Sheet on Hate Speech, *supra* note 139.

¹⁷⁶ See *id.*

¹⁷⁷ *Stomakhin v. Russia*, App. No. 52273/07, ¶¶ 90-91 (May 9, 2018), <http://hudoc.echr.coe.int/eng/?i=001-182731>.

interests of the state.¹⁷⁸ For example, in 2018, the European Court found Russia’s conviction of a journalist who appealed to “extremist activities through the mass media” and “incit[ed] hatred and enmity [and] humiliat[ed] the dignity of an individual or group” did not comport with Article 10 of the European Convention.¹⁷⁹ The Court found that certain of the journalist’s statements that called the Russian security forces “maniacs,” “murderers,” and other criminally-minded terminology, would “stir up a deep-seated and irrational hatred towards them in a clear attempt to justify and advocate violent actions against them,” and “expose them to the possible risk of physical violence.”¹⁸⁰ The European Court found that there was a pressing need and relevant and sufficient reasons to interfere with such speech.¹⁸¹ However, the Court found that the punishment (imprisonment for five years and a three-year prohibition on engaging in journalism) was disproportionate.¹⁸² In another application of its proportionality inquiry, the Court determined that Denmark’s conviction of a journalist for reporting on derogatory and racist comments was not proportionate to the aim of protecting the rights and reputations of others because the purpose of the speaker’s work was to expose and analyze racist groups, not to promote their racist message.¹⁸³

Within its proportionality analysis, the Court has often found that interferences with expression “capable” of harm are proportionate to the state’s legitimate aim. Further, it has done so without requiring the government to demonstrate with specificity the speaker’s intent to cause harm and the likelihood of near-term harm (i.e., imminence). For example, the European Court upheld Russia’s conviction of a journalist for inciting hatred because his article “could be reasonably assessed” as stirring up prejudices towards non-Russians.¹⁸⁴ Similarly, the European Court concluded that France’s conviction of a cartoonist for publishing satirical drawings after 9/11 was proportionate because those drawings could stoke violence and have a “plausible impact on public order.”¹⁸⁵ The Court has

¹⁷⁸ *Id.* ¶ 96; *Gündüz v. Turkey*, 2003-XI Eur. Ct. H.R. ¶ 49.

¹⁷⁹ *Stomakhin v. Russia*, App. No. 52273/07, ¶¶ 45, 133-34 (May 9, 2018).

¹⁸⁰ *Id.* ¶ 107.

¹⁸¹ *Id.* ¶ 109.

¹⁸² *Id.* ¶¶ 128-30.

¹⁸³ *Jersild*, App. No. 15890/89, ¶¶ 33-37 (Sept. 23, 1994).

¹⁸⁴ *Atamanchuk*, App. No. 4493/11, ¶ 64 (Feb. 11, 2020).

¹⁸⁵ *Leroy v. France*, App. No. 36109/03, ¶ 45 (Oct. 2, 2008), <http://hudoc.echr.coe.int/eng?i=001-88657>.

not required a finding of intent to incite near-term harm or likely near-term harm when upholding a variety of other hate speech convictions.¹⁸⁶

In its hate speech cases, the European Court has even explicitly stated that the intent of the speaker and the likelihood of imminent harm (which are key factors in the U.N.'s necessity approach) are not needed to uphold hate speech convictions.¹⁸⁷ When hate speech is not linked to potential violence or crime, the Court's decisions are "based on an assessment of the content of the expression and the manner of its delivery."¹⁸⁸ For example, the European Court has noted that "wanton denigration" and insulting speech is enough to place speech outside the protection of the European Convention.¹⁸⁹ In a 2020 case, the European Court noted that "attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for allowing the authorities to favour combating prejudicial speech within the context of permitted restrictions on freedom of expression."¹⁹⁰

C. *The African Human Rights System*

The central document in the African human rights system is the African Charter on Human and Peoples' Rights, which entered into force in 1986 and has 54 state parties.¹⁹¹ Article 9 of the African Charter provides that every person has the right to (1) "receive information," and (2) "to express and disseminate his opinions within the law."¹⁹² Though the African Court of Human and Peoples' Rights (African Court) has decided very few freedom of expression cases to date, some important hate speech observations can be drawn from its existing jurisprudence, as well as a recently adopted declaration.

In a case in which it overturned criminal sanctions for defamation, the African Court acknowledged that, unlike other international and regional

¹⁸⁶ See, e.g., *Nix v. Germany*, App. No. 35285/16, (Mar. 13, 2018), <http://hudoc.echr.coe.int/eng?i=001-182241>; *Soulas and Others*, App. No. 15948/03 (July 10, 2008); *Šimunić v. Croatia*, App. No. 20373/17 (Jan. 22, 2019), <http://hudoc.echr.coe.int/eng?i=001-189769>; *I.A. v. Turkey*, 2005-VIII Eur. Ct. H.R. 249; *Gündüz*, 2003-XI Eur. Ct. H.R. 257; *Vejdeland*, App. No. 1813/07 (Feb. 9, 2012); *Sürek v. Turkey* (No. 3), 1999-VI Eur. Ct. H.R. 353.

¹⁸⁷ See *supra* notes 184-186 and accompanying text (describing ECtHR views about hate speech that is not protected despite a lack of imminent harm or a speaker's intention to create harm).

¹⁸⁸ *Lilliendahl*, App. No. 29297/18, ¶ 36 (May 12, 2020).

¹⁸⁹ *Savva Terentyev*, App. No. 10692/09, ¶ 68 (Aug. 28, 2018); *Gündüz*, 2003-XI Eur. Ct. H.R., ¶ 41.

¹⁹⁰ *Lilliendahl*, App. No. 29297/18, ¶ 36 (May 12, 2020).

¹⁹¹ African Charter on Human and Peoples' Rights, June 27, 1981, 1520 U.N.T.S. 217, 21 I.L.M. 58, <https://www.achpr.org/legalinstruments/detail?id=49>; Ratification Table – African Charter on Human and Peoples' Rights, <https://www.achpr.org/ratificationtable?id=49> (last viewed Feb. 2, 2022).

¹⁹² African Charter on Human and Peoples' Rights, *supra* note 191, art. 9.

treaties, the African Charter does not explicitly provide for application of the legality, legitimacy, and necessity principles when governments limit speech.¹⁹³ However, the African Court has read into the text the need for application of the legality, legitimacy, and necessity tests in any restriction on freedom of expression in a way that displays the goal of convergence with U.N. standards.¹⁹⁴ For example, in the *Konaté* decision, the Court interpreted “within the law” in Article 9 to mean the law restricting speech must not be improperly vague.¹⁹⁵ Regarding the legitimacy test, the Court viewed the interests set forth in Article 27(2) of the African Charter (i.e., “the rights of others, collective security, morality and common interest”), as well as the public interest objectives in ICCPR Article 19(3), as constituting the legitimate reasons for which government can restrict speech.¹⁹⁶ With respect to the “necessity” condition, the African Court focused primarily on the need for restrictions to be proportionate.¹⁹⁷ Noting that the European Court viewed hate speech as warranting criminal sanctions, the African Court observed that it would view “incitement to international crimes” and “public incitement to hatred, discrimination or violence or threats” as warranting imprisonment.¹⁹⁸ The African Court did not include an analysis of the “least intrusive means test” in its decision, but instead turned to the European Court’s approach toward hate speech, which risks taking African jurisprudence down a path that departs from U.N. standards.

In a subsequent decision, *Ingabire Victoire Umuhoza v. Rwanda*, an appeal of a criminal conviction for minimization of genocide, the African Court reinforced the approach to the legality, legitimacy, and necessity tests that it articulated in its *Konaté* decision.¹⁹⁹ In addition, the Court explicitly endorsed granting governments a “margin of appreciation” in assessing freedom of expression claims,²⁰⁰ despite the U.N. Human Rights

¹⁹³ *Konaté v. Burkina Faso*, No. 004/2013, Decision, African Court on Human and Peoples’ Rights [Afr. Ct. H.P.R.], ¶ 129 (Dec. 5, 2014), <https://www.african-court.org/en/images/Cases/Judgment/Judgment%20Appl.004-2013%20Lohe%20Issa%20Konate%20v%20Burkina%20Faso%20-English.pdf>.

¹⁹⁴ *Konaté*, [Afr. Ct. H.P.R.], ¶¶ 129-31, 134-35, and 139-66.

¹⁹⁵ *Id.* ¶¶ 129-131.

¹⁹⁶ *Id.* ¶¶ 134-135.

¹⁹⁷ *Id.* ¶¶ 139-166.

¹⁹⁸ *Id.* ¶¶ 158, 165.

¹⁹⁹ *Umuhoza v. Rwanda*, No. 003/2014, Decision, African Court on Human and Peoples’ Rights [Afr. Ct. H.P.R.], ¶¶ 120-63 (Nov. 24, 2017), <https://www.african-court.org/cpmt/storage/app/uploads/public/5fa/a76/468/5faa7646871e5498448857.pdf>.

²⁰⁰ *Id.* ¶ 138 (stating that in “considering the margin of appreciation that the Respondent State enjoys in defining and prohibiting some criminal acts in its domestic legislation, the Court is of the

Committee's rejection of this doctrine.²⁰¹ Though the Court ultimately found that the criminal sanction for genocide minimization in this particular case violated the African Charter's freedom of expression guarantee,²⁰² it did note that "[s]tatements that deny or minimize the magnitude or effects of the genocide or that unequivocally insinuate the same fall outside the domain of the legitimate exercise of the right to freedom of expression and should be prohibited by law."²⁰³ Such an approach contrasts with the U.N.'s Human Rights and the ICERD Committees' approach to genocide denial,²⁰⁴ which were not cited in the opinion.

Following these opinions, in 2019, the African Commission on Human and Peoples' Rights adopted the Declaration of Principles on Freedom of Expression and Access to Information in Africa (African Declaration),²⁰⁵ which brings the African approach closer to U.N. standards on hate speech. For example, Principle 9(1) of the Declaration closely mirrors the U.N. interpretations of ICCPR Article 19's legality,²⁰⁶ legitimacy,²⁰⁷ and necessity tests.²⁰⁸ Notably, the African Declaration

view that the impugned laws provide adequate notice for individuals"). Thus the Court relied on the margin of appreciation doctrine in assessing the legality (i.e., vagueness) challenge to the law. *Id.*

²⁰¹ See *supra* note 154 and accompanying text.

²⁰² Umuhoza, [Afr. Ct. H.P.R.], ¶ 163.

²⁰³ *Id.* ¶ 158.

²⁰⁴ See *supra* notes 99-102 (observing that both Committees take the position that atrocity denial should not be prohibited absent incitement to likely and imminent harm). The European Court has, however condoned bans on genocide denials. See *supra* note 148 and accompanying text.

²⁰⁵ African Commission on Human and Peoples' Rights, *Declaration of Principles on Freedom of Expression and Access to Information in Africa* (Nov. 2019), <https://www.achpr.org/legalinstruments/detail?id=69> [hereinafter African Declaration].

²⁰⁶ The African Declaration defines the legality test as meaning that any speech restrictions must be "clear, precise, accessible and foreseeable . . . overseen by an independent body in a manner that is not arbitrary or discriminatory . . . and effectively safeguards against abuse including through the provision of a right of appeal to independent and impartial courts." *Id.* principle 9(2). The U.N.'s human rights monitors interpret legality similarly. See *supra* notes 38-41 and accompanying text (describing the vagueness test and the need for independent court adjudication).

²⁰⁷ The African Declaration defines legitimate public interest aims as encompassing all the aims in ICCPR Article 19(3) with the exception of "morals." Compare *African Declaration*, *supra* note 205, Principle 9(3), with the public interest objectives set forth in ICCPR Article 19(3), *supra* note 10 and accompanying text. In addition, unlike the African Court's prior decisions, the African Declaration did not endorse as "legitimate" objectives for speech restrictions the interests set forth in African Charter Article 27(2). Compare *African Declaration*, *supra* note 205, Principle 9(3), with the African Court's decisions in Konaté, [Afr. Ct. H.P.R.], ¶¶ 129-31, and Umuhoza, [Afr. Ct. H.P.R.], ¶ 140. However, an African Commission on Human Rights decision that was adopted the same month as the Declaration seems to have accepted African Charter Article 27's additional list of governmental interests. See *Uwimana-Nkusi v. Rwanda*, Communication 426/12, African Commission on Human and Peoples' Rights [Afr. Comm'n H.P.R.], ¶ 161 (Nov. 10, 2019), https://www.achpr.org/public/Document/file/English/Decision%20on%20Comm%20426-12_ENG.pdf (noting Article 27 permits restrictions on rights in the African Charter, including freedom of expression).

limited the scope of permissible governmental objectives²⁰⁹ and endorsed not only a proportionality analysis but also a “least restrictive means” analysis as part of the necessity test.²¹⁰ In addition, the African Declaration adopted a mandatory ban on hate speech that is phrased similarly to ICCPR Article 20²¹¹ and endorsed U.N. interpretations about only criminalizing such speech as a last resort after assessing contextual factors endorsed by U.N. experts (including the intent of the speaker as well as the likelihood and imminence of violence, discrimination, or hostility).²¹² The Declaration was silent as to whether the African human rights machinery should defer to governments by invoking the margin of appreciation doctrine when

²⁰⁸ The African Declaration defines the necessity condition as requiring that a speech limitation: originate from a pressing and substantial need that is relevant and sufficient; have a direct and immediate connection to the expression and disclosure of information, and *be the least restrictive means of achieving* the stated aim; and be such that the benefit of protecting the stated interest outweighs the harm to the expression and disclosure of information, including with respect to the sanctions authorised.

African Declaration, *supra* note 205, principle 19(4) (emphasis added).

²⁰⁹ *Id.* principle 9(3) (stating that only ICCPR Article 19(3) objectives, other than “morals,” qualify as legitimate objectives and dropping African Charter Article 27’s list of governmental interests).

²¹⁰ Compare *supra* note 208 (describing the Declaration’s approach to the “necessity” test, which explicitly includes a “least restrictive means” test and proportionality assessment) with *supra* notes 46-55 and accompanying text (describing how the U.N. system’s necessity test includes both a “least intrusive means” and a proportionality analysis as well as noting the requirement for a “direct and immediate connection” between the restriction and the expression).

²¹¹ *African Declaration*, *supra* note 205, principle 23(1) (“States shall prohibit any speech that advocates for national, racial, religious or other forms of discriminatory hatred which constitutes incitement to discrimination, hostility or violence”).

²¹² *Id.*, principle 23(2). When determining if hate speech rises to the level of severity for the imposition of criminal sanctions, states should consider

(a.) prevailing social and political context; (b.) status of the speaker in relation to the audience; (c.) *existence of a clear intent to incite*; (d.) content and form of the speech; (e.) extent of the speech, including its public nature, size of audience and means of dissemination; (f.) *real likelihood and imminence of harm*.

Id. (emphasis added). These factors are very similar to those in the U.N.’s Rabat Plan of Action as well as those endorsed by the U.N. Special Rapporteur in 2012. See *supra* notes 59-74 and accompanying text. In 2019, the African Commission on Human Rights issued a noteworthy decision that overturned Rwandan criminal convictions of journalists who had published articles critical of government policy (particularly of the president) and of the Gacaca court system set up following the 1994 genocide in the country. Uwimana-Nkusi, [Afr. Comm’n H.P.R.], (Nov. 10, 2019). Rwanda defended its criminal sanctions, *inter alia*, on the grounds that “chaos and [a] circle of violence” could result from such expression. *Id.* ¶¶ 75-80. While the Commission acknowledged that the history of the Rwandan genocide could counsel in favor of certain restrictions on expression, it found that the specific context involved—after reviewing the necessity standards (including a lack of intent and immediacy of consequences)—meant that the Rwandan convictions were inconsistent with Article 9 of the African Charter. *Id.* ¶¶ 175-209. However, the Commission also noted that “following the conclusion of the European Court that denial of Holocaust is not protected under freedom of expression, the Commission holds that in Rwanda as well expressions that entail denial of the genocide against the Tutsi cannot be protected under Art. 9 of the African Charter.” *Id.* ¶ 207. Such an approach departs from the U.N. Human Rights Committee’s view that atrocity denial is protected speech. See *supra* notes 99-102 and accompanying text.

adjudicating allegations of freedom of expression violations. While it will be important to monitor how the African Declaration is interpreted and implemented in freedom of expression cases by the African Court and Commission,²¹³ the Declaration itself provides a strong foundation for convergence with the U.N.'s approach to the legality, legitimacy, and necessity tests.

D. Emergent Regional Approaches

While the Americas, Europe, and Africa enjoy the most developed human rights systems, other regional and religious-based organizations have adopted approaches that seem to assert independence from U.N. legal standards in their normative frameworks. This Section begins by discussing the approach of the League of Arab States and the Association of Southeast Asian Nations (ASEAN), before turning to the OIC's religion-based approach. These newer and weakly-institutionalized initiatives do not have functioning and legally binding enforcement mechanisms like the Inter-American, European and African systems. In addition, the OIC and ASEAN approaches are centered on non-binding declarations rather than treaties.²¹⁴ This Section discusses these newer initiatives to highlight how their emerging normative approaches diverge significantly from global standards, signaling potential further challenges to the U.N. system.

The Arab League adopted the Arab Charter on Human Rights in 2004.²¹⁵ Article 32 of the Arab Charter sets forth the right to freedom of information and expression: "The present Charter guarantees the right to information and to freedom of opinion and expression, as well as the right

²¹³ In a case decided after the African Declaration was adopted, the complainant argued that enforcement of Benin's law that "punish[ed] the offences of racially motivated and xenophobic insults using a computer system and that of incitement to hatred and violence on such grounds as race, colour, national or ethnic origin, or religion" violated freedom of expression. *Ajavon v. Republic of Benin*, No. 062/2019, Decision, African Court of Human and Peoples' Rights [Afr. Ct. H.P.R.], ¶ 121 (Dec. 4, 2020), <https://www.african-court.org/cpmt/storage/app/uploads/public/602/10a/50c/60210a50cb6e2538353022.pdf>. In finding no freedom of expression violation, the Court did not set forth its analysis in detail. In applying the legality test, the Court did not commemorate an analysis of why the legal prohibition met the vagueness test nor discuss any national court interpretations that may have remedied any ambiguities; rather, it noted the test was met. *Id.* ¶ 122. In its necessity and proportionality analysis, the Court did not assess the speakers' intent or the likelihood of imminent harm in holding this test was met; rather, it asserted that possible harm was sufficient to meet this test. *Id.* ¶ 127.

²¹⁴ See *infra* notes 219 and 224 and accompanying text for a discussion of the relevant ASEAN and OIC declarations.

²¹⁵ League of Arab States, Arab Charter on Human Rights, May 22, 2004, *reprinted in* 12 Int'l Hum. Rts. Rep. 893 (2005) [hereinafter Arab Charter]. The Arab League has adopted a statute for an Arab Human Rights Court, but the Court is not yet functioning. *League of Arab States*, INTERNAT'L CTR NOT-FOR-PROFIT LAW (Sept. 25, 2021), https://www.icnl.org/wp-content/uploads/Transnational_arabcharter2004en.pdf (last visited Dec. 21, 2021).

to seek, receive and impart information and ideas through any medium, regardless of geographical boundaries.”²¹⁶ The Charter also states that: “Such rights and freedoms shall be exercised *in conformity with the fundamental values of society* and . . . subject only to such limitations as are *required* to ensure respect for the rights or reputation of others or the protection of national security, public order and public health or morals.”²¹⁷ The Charter differs from U.N. standards on freedom of expression in various ways. First, it does not mention that restrictions on speech must be “provided by law” to achieve legitimate governmental objectives (i.e., it omits the legality condition). Second, it states that individuals are expected to exercise their speech rights in accordance with the fundamental values of society (an amorphous concept that limits the speaker rather than the government), thereby presuming a “legitimate” public interest objective for speech limitations. And third, it is unclear how the word “required” will be interpreted. Will it include the concepts of least intrusive means and proportionality, as defined by the U.N.’s “necessary” test? These are three significant ways in which the Arab Charter risks departing from the U.N.’s rigorous tests of legality, legitimacy, and necessity. The failure of the Arab League to clarify its consistency with international human rights norms does not bode well in constraining problematic government limitations on speech.

While not a treaty with legally binding obligations, the 2012 ASEAN Declaration of Human Rights²¹⁸ also fell short of global human rights standards.²¹⁹ With respect to freedom of expression, the ASEAN Declaration provides a general statement of the right: “Every person has the right to freedom of opinion and expression, including freedom to hold opinions without interference and to seek, receive and impart information, whether orally, in writing or through any other medium of that person’s choice.”²²⁰ This is similar to the U.N. definition; however, the ASEAN Declaration contains a variety of limitations clauses, which temper all of the listed rights, thus appearing to depart from global standards. For example, Article 7 of the Declaration appears to espouse cultural relativism

²¹⁶ *Id.* art. 32.

²¹⁷ *Id.* (emphasis added).

²¹⁸ Ass’n of Southeast Asian Nations, ASEAN Human Rights Declaration, ¶ 23, Nov. 19, 2012, <https://asean.org/asean-human-rights-declaration> [hereinafter ASEAN Declaration].

²¹⁹ See U.N. Official Welcomes ASEAN Commitment to Human Rights, But Concerned Over Wording, U.N. NEWS (Nov. 19, 2012), <https://news.un.org/en/story/2012/11/426012#.UPgVKGckSOI> (noting that, given inconsistencies with the U.N.’s minimum human rights protection, it is important that problematic language in the Declaration not “become a part of any binding regional human rights convention.”).

²²⁰ ASEAN Declaration, *supra* note 218, ¶ 23.

in assessing the scope of rights rather than cementing global minimum standards in respect for human rights.²²¹ It will be important to monitor future ASEAN interpretations to determine if this Declaration is applied in a manner consistent with U.N. standards.²²²

Although the OIC's Cairo Declaration on Human Rights in Islam,²²³ an instrument that does not contain legally binding obligations, emerged from a religion-based (rather than geography-based) institution, it bears mentioning as the OIC is the second largest intergovernmental organization after the UN²²⁴ and is active in human rights norm-setting discussions in U.N. fora.²²⁵ The Cairo Declaration, which was revised and (re-)adopted in 2020,²²⁶ departs from U.N. standards on freedom of expression in a variety of ways.²²⁷ For example, it does not define the scope of the right and focuses instead on restrictions, which leaves the definition and scope of the

²²¹ *Id.* art. 7 (providing that “the realisation of human rights must be considered in the regional and national context bearing in mind different political, economic, legal, social, cultural, historical and religious backgrounds”).

²²² In 2019, the ASEAN Intergovernmental Commission on Human Rights hosted a consultation on freedom of expression and received “inputs on the formulation of the Recommendation on the Implementation of the Article 23 of the AHRD to serve as a common language for human rights cooperation and actions in the context of regionalism in ASEAN.” See *2019 AICHR Consultation on Freedom of Opinion and Expression in ASEAN (Article 23 of the ASEAN Human Rights Declaration)*, AICHR, <https://aichr.org/news/the-2019-aichr-consultation-on-freedom-of-opinion-and-expression-in-asean-article-23-of-the-asean-human-rights-declaration>. The eventual Recommendation on freedom of expression could serve as an opportunity to unequivocally align the ASEAN approach with the U.N.'s minimum universal standards.

²²³ Cairo Declaration, *supra* note 18.

²²⁴ See *OIC History*, *supra* note 17.

²²⁵ See, e.g., Suzanne Nossel, *Advancing Human Rights in the U.N. System*, COUNCIL ON FOREIGN RELATIONS 15-16 (2012), <https://perma.cc/P2TC-L3NC> (describing the role of the OIC in U.N. developments on blasphemy, freedom of religion, and freedom of expression).

²²⁶ See *supra* notes 18 and 21.

²²⁷ The Cairo Declaration's provision on freedom of expression provides:

Everyone shall have the right to freedom of expression. The exercise of this right carries with it special duties and responsibilities. The State has the obligation to protect and facilitate the exercise of this right while also protecting its legitimate national integrity and interests, as well as *promoting harmony, welfare, justice and equity* within society. Any restrictions on the exercise of this right, to be clearly defined in the law, and shall be limited to the following categories: i. Propaganda for war. ii. Advocacy of hatred, discrimination or violence on grounds of religion, belief, national origin, race, ethnicity, color, language, sex or socio-economic status. iii. Respect for the human rights or reputation of others. iv. Matters relating to national security and public order. v. Measures required for the protection of public health or morals.

The State and society shall endeavor to disseminate and promote the principles of tolerance, justice and peaceful coexistence among other noble principles and values, and to discourage hatred, prejudice, violence and terrorism. *Freedom of expression should not be used for denigration of religions and prophets or to violate the sanctities of religious symbols or to undermine the moral and ethical values of society.*

Cairo Declaration, *supra* note 18, art. 22(b)-(c) (emphasis added).

right unclear.²²⁸ Moreover, it departs from ICCPR Article 19(3)'s tripartite test for speech restrictions. It does not explicitly include a requirement that limitations on speech be "necessary," which risks omitting an analysis of whether speech restrictions constitute the least intrusive means and are proportional.²²⁹ In addition, the Cairo Declaration and related interpretations refer to grounds for restricting speech that go beyond those enumerated in the ICCPR,²³⁰ including by treating blasphemy as permissible grounds for restricting speech.²³¹ In adopting this approach, the OIC's Human Rights Commission praised the European Court of Human Rights' reluctance to prohibit the criminalization of blasphemy, speech that offends religious sensibilities, or displays religious intolerance.²³² In addition, Article 25 of the Cairo Declaration appears to limit all rights, including freedom of expression, based on religious interpretations and national legislation, which is inconsistent with international standards.²³³ It will be important to monitor OIC interpretations to see if such deficiencies in the Declaration's text are remedied, though a 2015 OIC Human Rights

²²⁸ Compare *id.* ("Everyone shall have the right to freedom of expression"), with ICCPR, *supra* note 10, 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").

²²⁹ See *supra* note 227 (setting forth the text of the Cairo Declaration's free expression provision, which does not contain a provision that speech limitations must be "necessary").

²³⁰ For example, the Cairo Declaration seems to declare as legitimate public interest objectives "national integrity and interests, as well as promoting harmony, welfare, justice and equity within society." See *supra* note 227. But the ICCPR does not contain such broad concepts. See ICCPR, *supra* note 10, at 19(3).

²³¹ The Cairo Declaration provides that "Freedom of expression should not be used for denigration of religions and prophets or to violate the sanctities of religious symbols or to undermine the moral and ethical values of society." *Supra* note 227. In addition, the revised Cairo Declaration contains "advocacy of hatred" as a legitimate basis for restricting speech. *Id.* A 2015 interpretation of the OIC Human Rights Commission took the position that "sheer disrespect, defamation, insult, and negative stereotyping constitute incitement to religious hatred," which should not be protected speech. *IPHRC 8th Session: Outcome Document of the Thematic Debate on 'Freedom of Expression' and 'Hate Speech'*, OIC-IPHRC 1, 1, https://oic-iphrc.org/en/data/docs/session_reports/8th/8th_iphrc_thematic_debate_outcome_en.pdf [hereinafter OIC 2015 Interpretation]. The U.N.'s human rights machinery, however, has observed that such goals do not constitute legitimate public interest objectives for restricting speech. See *supra* notes 103-05 and accompanying text.

²³² OIC-IPHRC, *supra* note 21 ("the Commission recalled and reaffirmed . . . the jurisprudence emerging from the European Court of Human Rights, which validates restrictions on freedom of expression criticizing religious beliefs where such expression constitutes incitement to hatred and is deemed offensive to the adherents of a particular religion.").

²³³ Cairo Declaration, *supra* note 18, art. 25 ("Everyone has the right to exercise and enjoy the rights and freedoms set out in the present declaration, without prejudice to the principles of Islam and national legislation").

Commission interpretation of freedom of expression emphasized fewer protections for speech than the U.N. system.²³⁴

III. REFLECTIONS & RECOMMENDATIONS

Part III(A) provides an overview of areas of convergence and conflict between U.N. and regional standards in the Inter-American, European, and African systems with respect to hate speech and freedom of expression. Part III(B) provides recommendations that human rights defenders can deploy to reinforce the U.N.'s minimum standards in litigation before national, regional, and international courts, as well as in advocacy efforts regarding domestic laws and the regulation of speech by private platforms.

A. *Comparison of Global and Regional Standards*

Both the U.N. and regional human rights systems allow for hate speech to be banned, but differences arise with the applicable criteria for judging whether hate speech bans are valid. These conflicts generally arise with respect to (1) governmental burdens to demonstrate the lawfulness of speech bans, and (2) the substantive content of the legality, legitimacy, and necessity conditions, which are foundational to assessing the lawfulness of speech bans.

Under the U.N. and Inter-American human rights systems, the government bears the burden of proving all restrictions on speech, including hate speech, are valid.²³⁵ This creates a presumption in favor of speech, with the government bearing the burden of proving the lawfulness of restrictions. In contrast, the European Court has espoused some jurisprudential doctrines that remove this burden.²³⁶ For example, the European Court considers certain hate speech to be so offensive that it deems freedom of expression-based appeals to be inadmissible, which relieves the government from justifying the lawfulness of its speech

²³⁴ In 2015, before the adoption of the most recent version of the Cairo Declaration, the OIC's Human Rights Commission adopted a variety of views and recommendations about hate speech and freedom of expression, many of which did not align with the U.N.'s minimum standards. OIC 2015 Interpretation, *supra* note 231. For example, the Outcome Document indicated that "sheer disrespect, defamation, insult, and negative stereotyping constitute incitement to 'religious hatred,'" which should not be protected speech. *Id.* at 1. It emphasized that the "need to protect the sanctity of religions and their symbols" (i.e., blasphemous speech) was to avoid the "defamatory stereotyping and insults" of believers who are then subject to discrimination, hostility, and violence. *Id.* at 2. It criticized the U.N. special procedures' approach to hate speech as unbalanced. *Id.* at 2. It cited favorably to the ICERD Committee's 1993 General Recommendation XV on Article 4 (regarding racist hate speech) rather than its 2013 views General Recommendation 35, which provides a more speech protective approach to combating racism (*see supra* notes 78-92).

²³⁵ *See supra* notes 30, 120 and accompanying text.

²³⁶ *See supra* notes 142-58.

restriction and leaves the national punishment for the expression in place.²³⁷ In addition, the European Court consistently uses its margin of appreciation doctrine to defer to judgments of governments, further lessening the state's burden to demonstrate compliance with free expression standards.²³⁸ The African Court has also invoked the European-created margin of appreciation doctrine in its jurisprudence.²³⁹ In a sharp, if implicit, rebuke to the European Court over a decade ago, the U.N. Human Rights Committee rejected the margin of appreciation doctrine for freedom of expression.²⁴⁰

With respect to the legality component of the three-part test for judging the lawfulness of speech bans, both the U.N. and Inter-American systems rigorously assess whether hate speech laws are improperly vague.²⁴¹ The European Court's jurisprudence, however, has displayed reluctance in acknowledging vagueness in hate speech laws.²⁴² Perhaps because of the European Court's approach to vagueness, litigants frequently concede this issue.²⁴³ The African Court has interpreted the African Convention as encompassing a requirement that laws must not be improperly vague, though its hate speech jurisprudence is in too early a phase to assess how rigorously it will apply this standard.²⁴⁴

With regard to the legitimacy test, both the U.N. and Inter-American systems contain similar public interest objectives that can be invoked to limit speech.²⁴⁵ Both systems also take the position that protecting religions from criticism or protecting adherents of religions from offense are not legitimate public interest objectives.²⁴⁶ The European Convention contains a longer list of public interest objectives, and the European Court has opined that the protection of religious sensibilities is an appropriate objective for limiting speech.²⁴⁷ In addition, the U.N. human rights system takes the position that speech that denies historic facts (when untethered to incitement to particular harms) is permissible, whereas the European Court has determined that the denial of certain historic atrocities is so grave that such cases are outside the protection of the European Convention's

²³⁷ See *supra* notes 142-48.

²³⁸ See *supra* notes 150-58.

²³⁹ See *supra* note 200 and accompanying text.

²⁴⁰ See *supra* note 154 and accompanying text.

²⁴¹ See *supra* notes 38-41, 94-98, 121 and accompanying text.

²⁴² See *supra* notes 159-171 and accompanying text.

²⁴³ See *supra* note 171 and accompanying text.

²⁴⁴ See *supra* note 195 and accompanying text.

²⁴⁵ See *supra* notes 42-45, 119 and accompanying text.

²⁴⁶ See *supra* notes 103-04, 172 and accompanying text.

²⁴⁷ See *supra* notes 141, 172 and accompanying text.

protection for freedom of expression.²⁴⁸ The African Court has recognized a list of public interest objectives for limiting expression that is broader than those set forth in ICCPR Article 19(3), but the African Declaration has espoused a more restrictive view of legitimate public interest aims.²⁴⁹

The U.N. and Inter-American systems both assess if a speech restriction is “necessary” by, among other things, evaluating if it (1) is the least intrusive means to achieve the legitimate public interest objective, and (2) is proportional to the aim to be achieved.²⁵⁰ Both the U.N. and Inter-American systems have stressed the importance of using non-censorial methods of tackling intolerance prior to resorting to speech bans.²⁵¹ In assessing the necessity of speech restrictions, the U.N. system has emphasized the importance of governments demonstrating likely and imminent harm²⁵² as well as the need to display the speaker intended to cause harm.²⁵³ The European Court does not apply a least intrusive means test, but rather examines restrictions under a proportionality test and has allowed speech bans where governments have not demonstrated likely or imminent harm.²⁵⁴ In addition, whereas the U.N.’s proportionality test is concerned with assessing if the infringement on the expression rights of the speaker and of others is outweighed by the legitimate aim, the European Court appears to allow for a more amorphous weighing of various rights and interests.²⁵⁵ Early cases from the African Court have focused primarily on proportionality rather than the least intrusive means test, though the African Declaration has embraced both a “least restrictive means” test as well as a proportionality assessment in determining if speech restrictions are necessary.²⁵⁶

While the U.N.’s ICCPR (Article 20), ICERD (Article 4), and the American Convention (Article 13.5), contain mandatory bans on certain hate speech, the European Convention and African Charter do not contain such bans (though the African Declaration did espouse a mandatory ban

²⁴⁸ See *supra* notes 99-102, 148 and accompanying text.

²⁴⁹ See *supra* notes 196, 207 and accompanying text.

²⁵⁰ See *supra* notes 46-55, 124-25 and accompanying text.

²⁵¹ See *supra* notes 52-53, 135-36 and accompanying text.

²⁵² See *supra* note 108 and accompanying text.

²⁵³ See *supra* note 110 and accompanying text.

²⁵⁴ See *supra* notes 174-90 and accompanying text.

²⁵⁵ See *supra* notes 54-55, 177-90 and accompanying text. See also David A. Kaye, *Against Balancing* (May 4, 2020), [https://dkisaway.medium.com/against-balancing-5ecc5dc2beb5_\(explaining-the-dangers-of-balancing-freedom-of-expression-against-other-rights-rather-than-engaging-in-a-strict-application-of-the-legality-legitimacy-and-necessity-tests\)](https://dkisaway.medium.com/against-balancing-5ecc5dc2beb5_(explaining-the-dangers-of-balancing-freedom-of-expression-against-other-rights-rather-than-engaging-in-a-strict-application-of-the-legality-legitimacy-and-necessity-tests)).

²⁵⁶ See *supra* notes 197, 208-10 and accompanying text.

similar to ICCPR Article 20).²⁵⁷ Although the European Convention does not contain a mandatory hate speech ban, the European Court appears to apply lower thresholds for the prohibition of hate speech than the U.N. and Inter-American systems.²⁵⁸ For example, the U.N. and Inter-American systems both require that imposition of their mandatory hate speech bans include a finding of specific intent to cause particular harm, and that such harm be likely and imminent.²⁵⁹ However, the Inter-American Convention's ban covers a narrower range of harms than those set forth in ICCPR Article 20 or ICERD Article 4.²⁶⁰

This mapping exercise has not only illuminated key areas of divergence in which regional systems provide fewer rights than the U.N. system, but also revealed the potential for the European Court's prolific jurisprudence to (adversely) influence evolving jurisprudence in other regions. For example, the African system has cited favorably to a variety of European Court views – including on the margin of appreciation, the test for incitement, and the banning of atrocity denial²⁶¹ – that improperly depart from U.N. standards. Similarly, the OIC Human Rights Commission invoked the European Court's jurisprudence to justify its departure from U.N. freedom of expression protections with respect to blasphemy, speech that offends religious sensibilities, and expression that displays religious intolerance.²⁶²

B. Recommendations for Human Rights Defenders

The fact that some regional human rights bodies – particularly in the European system – are developing approaches that provide fewer protections for freedom of expression does not throw into disarray the substance of U.N. standards. Rather, countries violate their U.N. treaty obligations when they provide fewer protections than their international treaty obligations mandate.²⁶³ Such normative divergence between regional and U.N. approaches can result in confusion with respect to international standards and even facilitate challenging U.N. standards based on

²⁵⁷ See *supra* notes 56, 76, 127, 211 and accompanying text. The fact that the African Declaration has espoused the approach of ICCPR Article 20 indicates a commitment to aligning with international human rights standards.

²⁵⁸ See *supra* notes 56-92, 128-133, 184-190 and accompanying text.

²⁵⁹ See *supra* notes 56-92, 128-133 and accompanying text.

²⁶⁰ See *supra* notes 64-65, 84-85, 132-33 and accompanying text.

²⁶¹ See, e.g., *supra* notes 198, 200-201, 212 and accompanying text (describing instances in which the African system cited favorably to the European Court's approach, which differs from U.N. standards).

²⁶² See *supra* notes 230-33 and accompanying text.

²⁶³ See *supra* note 22 and accompanying text.

arguments animated by cultural relativism.²⁶⁴ Though such arguments may appear more principled because they are grounded in regional approaches, they are contrary to international law and, if unaddressed, can set the stage for erosion of U.N. minimum standards for the protection of freedom of expression.

Human rights defenders should therefore be alert to the potential for misuse of regional norms to weaken, or avoid implementation of, U.N. standards in a variety of settings. For example, governments may improperly invoke regional norms – particularly those derived from the European Court’s jurisprudence – to define the full scope of their international obligations in developing national legislation or defending existing laws before domestic and regional courts. In addition, countries are likely to continue invoking regional approaches to rationalize their U.N. treaty violations when questioned by U.N. monitors²⁶⁵ and when developing regional human rights instruments and interpretations.²⁶⁶ Moreover, governments will likely seek ways to modify universal norms by codifying regional approaches in relevant state-adopted U.N. human rights instruments²⁶⁷ as well as in their efforts to influence U.N. expert interpretations on the scope of ICCPR obligations.²⁶⁸

To the extent that human rights defenders face arguments in various fora that adhering to U.N. standards is problematic because they “disregard” the cultural and local conditions embodied in regional approaches, human rights defenders should raise two arguments. First, the U.N.’s legality, legitimacy, and necessity conditions do not force the homogenization of all speech laws. Rather the three-part test is a safety check to ensure that individuals are not being judged under laws that do not give appropriate notice (legality), that laws are imposed for important public interest purposes (legitimacy), and that governments do not use hammers to regulate speech when tweezers are sufficient (necessity).

²⁶⁴ See *supra* note 23 and accompanying text.

²⁶⁵ See, e.g., *supra* note 16 and accompanying text (describing Germany’s response to a U.N. inquiry by invoking the European Court’s approach to hate speech to justify its NetzDG law).

²⁶⁶ See, e.g., *supra* notes 19-21 and accompanying text (highlighting how the OIC’s Human Rights Commission invoked the European Court’s jurisprudence to justify the OIC Declaration’s departure from U.N. standards on speech that offends religious sensibilities).

²⁶⁷ See, e.g., Nossel, *supra* note 225, at 10-11 (describing the OIC’s effort during a decade at the United Nations to develop norms that commemorated its views on blasphemy).

²⁶⁸ For example, when the Human Rights Committee requested comments on its draft General Comment on freedom of assembly, Germany submitted recommendations to restrict the interpretations in the draft by citing to European Court cases. See U.N. Human Rights Committee, *Call for Comment: No. 37 on Article 21 of the International Covenant on Civil and Political Rights – Right of Peaceful Assembly*, OHCHR (Sept. 17, 2020), <https://www.ohchr.org/EN/HRBodies/CCPR/Pages/GCArticle21.aspx> (click “Inputs Received” dropdown, then select “Germany”).

Second, the U.N. approach does in fact require a robust examination of local conditions and context in order to assess the validity of speech restrictions.²⁶⁹ It is impossible to examine whether a restriction constitutes the least intrusive means or is proportional without understanding the local context. Similarly, it is not feasible to assess a speaker's intent or whether speech could lead to incitement to imminent and likely harm without understanding local context.

In order to counter misuse of regional norms to redefine (and weaken) U.N. standards in various fora, human rights defenders should remain vigilant with respect to (1) improper lessening or removal of governmental burdens to demonstrate the validity of speech restrictions, (2) applications of watered-down versions of the U.N.'s legality, legitimacy, and necessity conditions, and (3) neglect of the high thresholds contained in the U.N.'s mandatory speech bans. For instance, if governments claim they are entitled to a margin of appreciation when restricting speech, or assert that certain speech falls outside the protection of human rights law, human rights defenders should make clear that such regional notions directly conflict with existing U.N. standards.²⁷⁰

With respect to the legality, legitimacy, and necessity conditions, human rights defenders should advocate for implementation of U.N. standards relating to this three-part test and not acquiesce to regional institutions' different (and weaker) versions of these three conditions. Whether assessing draft legislation or arguing in a variety of court settings, human rights defenders should hold governments to the U.N.'s rigorous regime for assessing vagueness of prohibitions rather than accept, for example, the European system's high tolerance for vague prohibitions.²⁷¹ In their advocacy, human rights defenders should also argue for the U.N.'s restrictive view of legitimate public interest objectives (which do not include protection from religious offense or denial of historic facts).²⁷² Moreover, human rights defenders should insist that governments demonstrate that their speech restrictions meet the U.N. necessity test's least intrusive means analysis as well as its proportionality examination, rather than going along with the European Court's amorphous proportionality test.²⁷³ In sum, while the "name" of the three-part test is the same across the U.N. and regional systems (i.e., the legality, legitimacy,

²⁶⁹ See, e.g., *supra* notes 71-85 and accompanying text (describing how U.N. standards on hate speech and freedom of expression are tied to a contextual analysis).

²⁷⁰ See *supra* notes 235-40 and accompanying text.

²⁷¹ See *supra* notes 241-43 and accompanying text.

²⁷² See *supra* notes 245-49 and accompanying text.

²⁷³ See *supra* notes 250-56 and accompanying text.

and necessity test), the substantive content that is applied can differ, and human rights defenders should consistently argue for application of the U.N.'s version of each prong of the test.

When governments seek to justify laws based on the mandatory bans in ICCPR Article 20 and/or ICERD Article 4, human rights defenders should proactively remind governments and adjudicators that such laws must not only survive scrutiny under ICCPR Article 19(3)'s tripartite test, but also respect the strict thresholds delineated by U.N. experts.²⁷⁴ First, specific intent on the part of the speaker to incite harm must be demonstrated.²⁷⁵ Negligence or reckless disregard for potential harm are not sufficient to rise to the level of "advocacy" by a speaker. Second, the potential harm must not be speculative, but rather, the imminence and likelihood of the harm are key aspects to the U.N.'s assessment of "incitement."²⁷⁶ As cautioned by the U.N. Special Rapporteur on Freedom of Expression, mere "*advocacy of hatred* on the basis of national, racial or religious grounds is not an offence in itself."²⁷⁷

With regard to the particular harms that are embodied in the U.N.'s mandatory hate speech bans, human rights defenders should advocate for use of definitions proposed by U.N. experts. While ICCPR Article 20's harms of "discrimination" and "violence" may be well understood, "hostility" can pose interpretative issues due to its ambiguity and potential breadth. Because of this, human rights defenders should emphasize to states that they should interpret "hostility" in ICCPR Article 20 to require "a manifestation of hatred beyond a mere state of mind,"²⁷⁸ which can reasonably be understood to include lawless acts such as trespass and vandalism. A manifestation of hatred could also take the form of what the U.S. legal system calls a "true threat," which "encompass[es] those statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals."²⁷⁹

While remaining attentive to misuse of regional norms to justify departures from (and the potential weakening of) U.N. standards, human rights defenders should also emphasize the importance of implementing

²⁷⁴ See *supra* notes 56-89 and accompanying text.

²⁷⁵ See *supra* notes 61, 80 and accompanying text.

²⁷⁶ See *supra* notes 62, 82-83 and accompanying text.

²⁷⁷ *Special Rapporteur 2012 Report to UNGA*, *supra* note 60, ¶ 43. Unfortunately, regional approaches have at times departed from such thresholds. See, e.g., *supra* notes 198, 232 and accompanying text (noting instances when the African and OIC mechanisms have cited favorably to ECtHR approval of bans on mere "incitement to hatred" or "advocacy of hatred").

²⁷⁸ *Special Rapporteur 2012 Report to UNGA*, *supra* note 60, ¶ 44(e).

²⁷⁹ *Virginia v. Black*, 538 U.S. 343, 359 (2003).

U.N. Human Rights Council Resolution 16/18, which affirmed that governments should seek to engage in a wide variety of good governance measures to tackle hate and intolerance before resorting to speech bans.²⁸⁰ Countries that do not engage in such good governance are not to be rewarded with wide censorial powers as an initial approach to tackling hate and intolerance.²⁸¹ Often this hard work of deploying effective measures to tackle hate and intolerance is ignored, with the debate around hate speech focusing solely on banning speech as the means to tackle hate speech. Focusing exclusively on speech restrictions is inconsistent with the due diligence governments and courts should undertake in verifying that governments are implementing the least intrusive means of tackling hate.

Human rights defenders should also seek to engage with the Arab League, ASEAN, and the OIC to positively influence the problematic trajectory of normative developments in such fora. Within the last thirty years, each of these emerging approaches was formed by groups of states that included many who were bound by the ICCPR and yet which agreed to norms that, on their face, depart significantly from U.N. standards. For example, in 2020, the OIC – the second largest intergovernmental organization after the U.N. – adopted an updated Cairo Declaration that omitted key protections for freedom of expression and its human rights commission explicitly relied on European Court jurisprudence to justify its departure from U.N. standards.²⁸² Engaging with these newer mechanisms should be part of human rights defenders’ strategies for preserving the U.N.’s minimum standards on freedom of expression.

Finally, the divergence between U.N. and regional standards is relevant to the work of human rights defenders, given that private companies are seeking to align their platforms’ corporate speech rules with international human rights norms. Indeed, the world’s leading business and human rights framework – the U.N. Guiding Principles for Business and Human Rights (UNGPs) – calls on companies to align their operations with U.N. human rights standards.²⁸³ Some companies – such as Twitter and

²⁸⁰ See *supra* notes 52-53 and accompanying text.

²⁸¹ See *supra* notes 52-53 and accompanying text.

²⁸² See *supra* notes 224-34 and accompanying text.

²⁸³ Human Rights Council Res. 17/4, U.N. Doc. A/HRC/RES/17/4, ¶ 1 (July 6, 2011); see also John Ruggie (Special Representative of the Secretary General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises), *Rep. of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises: Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework*, U.N. Doc. A/HRC/17/31 (Mar. 21, 2011). UNGP Principle 12 defines “internationally recognized human rights” as including at a minimum the Universal Declaration of Human Rights and two foundational U.N. treaties (including the ICCPR) as well as an

Facebook – have adopted policies that state that they look to the ICCPR and regional treaties such as the European Convention on Human Rights.²⁸⁴ Such statements of policy should cause human rights defenders to raise a variety of questions in their engagement with online platforms that seek to align with the UNGPs. First, are the companies aware of the areas of conflict between the ICCPR and European approaches to freedom of expression and hate speech? If so, how are they applying different bodies of standards when there are evident areas of conflict? In other words, are they applying the U.N.’s approach on vagueness or that of the European Court? Are they applying the least intrusive means test or disregarding it, as does the European Court? Second, on what principled basis are companies selecting to elevate one region’s human rights instrument and disregarding those of other sub-global systems (or are they citing European instruments to appease active regulators)?²⁸⁵ Are the companies applying European standards to individuals in other regions? If so, what would be the principled basis for doing so? In sum, human rights defenders should urge private platforms seeking to implement the UNGPs to understand the areas of convergence and conflict between U.N. and regional human rights systems, and to align with the U.N.’s minimum standards for the protection of expression.

CONCLUSION

This Article has illuminated two countervailing and under-reported trends: on the one hand, a growing consensus among U.N. experts and treaty bodies concerning interpretations of hate speech prohibitions in international law; and on the other, a failure of several regional human rights bodies to develop approaches to hate speech that are consistent with

International Labor Organization declaration. *Id.* at 13. The Commentary for Principle 12 notes other U.N. human rights instruments may also be consulted. *Id.* at 14.

²⁸⁴ Twitter’s policies provide that its respect for freedom of expression is “grounded” in U.S. domestic law and the European Convention on Human Rights and “informed” by the UNGPs. *Defending and Respecting the Rights of People Using Our Service*, TWITTER, <https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice>. Facebook’s Corporate Human Rights Policy provides that, in seeking to respect human rights as provided for in the UNGPs, the company will turn to not only U.N. human rights instruments but also the Charter of Fundamental Rights of the European Union and the American Convention. FACEBOOK, CORPORATE HUMAN RIGHTS POLICY, *supra* note 8.

²⁸⁵ In the last few years, Europe has been actively seeking ways to regulate online hate speech. See David Kaye, *How Europe’s New Internet Laws Threaten Freedom of Expression: Recent Regulation Risk Censoring Legitimate Content*, FOREIGN AFFS. (Dec. 18, 2018), <https://www.foreignaffairs.com/articles/europe/2017-12-18/how-europes-new-internet-laws-threaten-freedom-expression> (describing a “wave” of European regulation affecting platforms and online speech). It may be that companies are citing to European human rights standards in an attempt to curry favor with European Union and national regulators. *See id.*

the U.N.'s universal standards. The U.N.'s approach requires that governments bear the burden of demonstrating that all speech restrictions pass the principled (and rigorously applied) test of legality, legitimacy, and necessity. In addition, the U.N.'s mandatory hate speech bans require establishing the speaker's specific intent and assessing whether harm is likely and imminent. While some regional approaches converge with the U.N.'s approach to freedom of expression and hate speech, others conflict by providing fewer protections for freedom of expression. Specifically, certain regional approaches (1) lessen the burden on governments to justify speech restrictions, (2) do not rigorously assess the vagueness of hate speech laws, (3) approve of illegitimate public interest objectives for restricting speech, and/or (4) neglect the least intrusive means test or provide a weaker proportionality test. The European system's divergence from U.N. norms is particularly noteworthy given its abundant jurisprudence and potential to affect the development of other regional approaches. To promote the convergence of regional systems and national laws with the U.N.'s approach to hate speech, human rights defenders across the globe should prioritize tackling this troubling trend of norm divergence to preserve universal minimum protections for freedom of expression.

