CGHR WORKING PAPER #7

The Right to be Free from the Harm of Hate Speech in International Human Rights Law

Mona Elbahtimy
Department of Politics and International Studies, University of Cambridge

Centre of Governance and Human Rights Working Papers

The **Centre of Governance and Human Rights (CGHR)**, launched in late 2009, draws together experts, practitioners and policymakers from the University of Cambridge and far beyond to think critically and innovatively about pressing governance and human rights issues throughout the world, with a special focus on Africa. The Centre aims to be a world-class interdisciplinary hub for fresh thinking, collaborative research and improving practice.

The CGHR Working Papers Series is a collection of papers, largely peer-reviewed, focussed on cross-disciplinary research on issues of governance and human rights. The series includes papers presented at the CGHR Research Group and occasional papers written by CGHR Associates related to the Centre's research projects. It also welcomes papers from further afield on topics related to the CGHR research agenda.

Series Editors: Sharath Srinivasan & Thomas Probert

Publisher: Centre of Governance and Human Rights, University of Cambridge

Contact: cghr@polis.cam.ac.uk, (+44) (0)1223 767 257

Abstract: *The current challenges posed by hate speech across the globe have prompted the need to better understand the evolution of the right to be free from the harm of hate speech as codified within Article 20(2) of the International Covenant on Civil and Political Rights. This paper examines the right's evolution within international human rights law (IHRL). Its purpose is to understand the core difficulties that have faced efforts geared to the development, strengthening and expansion of international standards that provide protection from the harm of hate speech. To elaborate upon such difficulties, the paper identifies four internal features of the right to be free from the harm of hate speech, representing the challenges facing its interpretation and implementation. These four features are the right's 'emotional' component; the complexities in proving the proscribed incitement; the tensions between the listeners' and speakers' rights to liberty and equality; and the right's group-identity component, which creates tensions between individual and group rights. The paper argues that these four internal features of the right have a strong and direct influence on understanding the difficult path the right has taken in its evolution within IHRL.*[*]

Both legal scholars and politicians have long been preoccupied with the challenge of reconciling the protection against hate speech with the safeguarding of fundamental freedoms, particularly freedom of expression. Nevertheless, the last few years have witnessed a new wave of public, political and academic debates regarding the basic questions and controversies underlying the legal regulation of hate speech. This regulation has taken on significant importance in the political agendas of many states across the globe, and has become among the most pressing, yet controversial, issues confronting international human rights law (IHRL).

A number of significant developments have shed new light on the challenges of hate speech and have provided a new context for its existence and dissemination. First and foremost, as the result of the unprecedented rise in immigration flows, most modern societies have increasingly become more diverse racially, religiously and culturally. In many cases, this diversity has accompanied social anxieties and inter-groups tensions, which provide fertile ground for hate speech and exacerbate its harms. In addition, the exponentially accelerating advances in information and communication technologies (ICT) have unintentionally provided a strengthened infrastructure for the proliferation of hate speech with increasing potency, speed and visibility. With the effects of the geographical dimension of globalization (the rising immigration flows) and its virtual dimension (the ICT revolution), certain hate speech incidents can now grow from mere national crises into global crises with wide-ranging and cross-boundary repercussions.

The contemporary features, manifestations and challenges of hate speech have drawn renewed attention to the international normative human rights framework on hate speech. Moreover, they have posed a number of questions: is a global answer in the law still needed, and is the international regulatory framework, as it currently stands, suitable to address the recent hate speech challenges? Or, on the other hand, does this framework need to be expanded in order to account for, and accurately reflect, the new globalized dynamics of hate speech.

Against the background of the hate speech challenges proliferating in a globalizing world and the questions they pose regarding the current international normative framework on hate speech, this paper sheds light on how the norm against hate speech has evolved in IHRL. Specifically, it examines the norm's evolution in three main aspects; its emergence in the ICCPR; the body of hate speech jurisprudence developed by supra-national monitoring and adjudicatory bodies; and the recent attempts in the UN to create new international standards on hate speech.

[*] This paper is based upon a PhD thesis "The right to be free from the harm of hate speech in international human rights law: An analysis of a difficult evolutionary path," submitted to the University of Cambridge in September 2013. The author would like to thank her supervisors Dr. Barbara Metzger and Dr. Sharath Srinivasan for their guidance throughout the process of writing the thesis as well as Dr. Thomas Probert for his editorial efforts on this paper.

This paper traces the norm's evolution through the abovementioned three domains and more importantly explains the core challenges involved in this regard. It argues that four internal features of the right are instrumental in accounting for the challenges that have faced the development, strengthening and expansion of international standards providing protection against the harm of hate speech. These four internal features of the right are the 'emotional' component; the complexities of proving incitement; the tensions between speakers' and listeners' rights to liberty and equality; and the group-identity component. Through the lens of these four intrinsic qualities of the right, and by illuminating the definitional complexities and tensions underlying these qualities, this paper argues that they provoke enduring controversies pertaining to the right's exact interpretation; furthermore, they limit its expansive potential.

I.

Article 20(2) of the ICCPR articulates the main and most comprehensive international standard that should guide states' legal regulation of hate speech. The narrow conception of Article 20(2) that only deals with it as a limitation of the exercise of freedom of expression, rather than a codification of an autonomous right, is unjustified. This Article sets forth the right to be free from incitement to discrimination, hostility or violence resulting from the advocacy of national, racial, or religious hatred. This paper refers to that right using the abbreviated term *'the right to be free from the harm of hate speech'* and argues that IHRL lexicon should recognize this right as such.

In fact, the question of what valid justifications for recognizing specific claims as human rights has been subject to extensive philosophical, legal and political debates and remains contested. Traditionally, appeals have been made to natural law and the inherent dignity of human beings.[1] Currently, an increasing number of theorists focus on grounding rights in human interests.[2] These theorists have suggested a wide range of interests, corresponding to different views of human life; however, there is a narrow margin of agreement among them on the valid criteria for determining when an interest deserves recognition as, or is sufficiently important to necessitate the formulation of, a human right.[3] These debates need not be repeated here, since this paper uses the term 'human rights' in a technical and positivist sense: as a legal term of art, referring to those rights that international human rights instruments have codified.[4] Through the political endorsement of states, these instruments validate and legitimize claims to protect specific interests and prevent certain harms. These instruments' provisions elevate such claims to the status of international human rights and create a universal vocabulary to describe the normative content of the rights they have codified.[5] While one might contest the theoretical or philosophical bases of international human rights instruments, these instruments remain to provide legal grounding for human rights protection.

---

[1] See Richard Tuck, *Natural Rights Theories: Their Origin and Development* (Cambridge: Cambridge University Press, 1979).

[2] John Tobin, *The Right to Health in International Law* (Oxford: Oxford University Press, 2012), 52.

[3] Alon Harel, "What Demands Are Rights? An Investigation into the Relation Between Rights and Reasons," *Oxford Journal of Legal Studies* 17, no. 1 (1997): 101–114; Tobin, *The Right to Health in International Law*, 51–53; James Griffin, *On Human Rights* (Oxford: Oxford University Press, 2008), 179–187; Amartya Sen, "Elements of a Theory of Human Rights," *Philosophy & Public Affairs* 32, no. 4 (2004): 320–325; Allen Buchanan, "The Egalitarianism of Human Rights," *Ethics* 120, no. 4 (July 2010): 120; Allen E Buchanan, *Justice and Health Care Selected Essays* (Oxford: Oxford University Press, 2009), 213, http://public.eblib.com/EBLPublic/Public View.do?ptiID=472216.

[4] David Feldman, *Civil Liberties and Human Rights in England and Wales*, 2nd ed. (Oxford: OUP, 2002), 5.

[5] Samantha Besson, "Human Rights: Ethical, Political or Legal? First Steps in a Legal Theory of Human Rights," in *The Role of Ethics in International Law*, ed. Donald Earl Childress (Cambridge; New York: Cambridge University Press, 2012), 237; Tobin, *The Right to Health in International Law*, 46, 54, 58, 73–74.

The ICCPR stands at 'the apex of human rights law'[6] as 'the most authoritative expression of the contemporary and universally accepted minimum standard of human rights.'[7] The Covenant's provisions proclaim and protect legal human rights as such, rather than important interests, concerns or aspirations.[8]States parties to the ICCPR are subject to legal obligations to respect, protect, and promote rights codified therein with regard to people within their jurisdiction. The present paper, through this positivist lens, recognizes the incorporation of Article 20(2) within the ICCPR as elevating the claim to be free from incitement to discrimination, hostility, or violence resulting from advocacy of hatred to the status of an international legal human right.

The distinct formulation of Article 20(2), when compared to other provisions of the ICCPR, as well as the fact that it takes effect by shrinking the available zone of freedom of expression, should not lead to an automatic presupposition that the Article does not set forth an independent right. On the contrary, it is the protection of an independent right that distinguishes the Article's formulation from other provisions within the ICCPR that allow states to impose interest-based limitations on rights without setting forth independent rights. According to those limitation provisions, states are granted discretionary power to apply limitations to the exercise of freedoms only as an option (i.e. these limitations are permissible, rather than mandatory, in nature). On the other hand, the mandatory nature of states' obligations under Article 20(2) corresponds to its right-declaratory nature, whereby the interest in being protected, in an absolute manner, against the harm of hate speech carves out an independent right. The absolute prohibitions of torture and slavery under Articles 7 and 8 have been widely acknowledged in the IHRL lexicon as duties upon states that give rise to the rights to be free from torture and slavery.[9] Similarly, the prohibition of incitement to discrimination, hostility, or violence is the duty that gives rise to the right to be free from the harm of hate speech (or acts as this right's counterpart obligation).

Furthermore, the autonomous presence of Article 20(2) in the Covenant's text corresponds to its right-declaratory nature. This contrasts with the Covenant's limitations provisions, which exist as sub-clauses within relevant articles. The free-standing status of Article 20(2) is also relevant to this nature, in that it indicates that once the Article's threshold is met, it not only restricts freedom of expression, but also other freedoms (such as freedoms of religion and assembly). Article 20(2) also entails a negative claim vis-à-vis the state, similar to other right-declaratory articles of the Covenant. Though it might appear prima facie as incorporating only a positive claim vis-à-vis the state to enact laws that prohibit the expressive acts described therein, Article 20(2) obliges the state to refrain from engaging in advocacy of hatred that constitutes incitement to discrimination, hostility, or violence. Indeed, the laws that should be enacted pursuant to Article 20(2) apply equally to private persons and state organs.[10]

As a consequence of the inclusion of Article 20(2) within the text of the ICCPR, the protection from incitement to discrimination, hostility or violence represents not only a societal or public interest that enhances the values of tolerance, mutual respect and dignity, but is also properly characterized as an international human right. The Article imposes rights-based, not merely interest-based, limitations on the exercise of freedoms. This distinction has clear practical consequences. While the violation of a right provides grounds (uncontested by theories of rights)

[6] Rosalyn Higgins, "The United Nations: Still a Force for Peace," *The Modern Law Review* 52, no. 1 (1989): 1.
[7] Manfred Nowak, *U.N. Covenant on Civil and Political Rights: CCPR commentary*, 2nd ed. (Kehl, Germany; Arlington, Va., USA: N.P. Engel, 2005), xi.
[8] Alon Harel, "What Demands Are Rights? An Investigation into the Relation Between Rights and Reasons," 113.
[9] Nowak, *U.N. Covenant on Civil and Political Rights*, 157,194.
[10] "Human Rights Committee, General Comment 11 on Article 20 (Nineteenth Session, 1983), Reprinted in Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies, U.N. Doc. HRI/GEN/1/Rev.6 at 133 (2003).," n.d., para. 2; Nowak, *U.N. Covenant on Civil and Political Rights*, 475.

for restricting the exercise of liberties, the protection of public or societal interests does not necessarily enjoy the same legitimate status. Moreover, at the adjudication level, the protection of public interests is not always prioritised over the protection of rights; precedence is granted to the right unless there is a compelling reason or danger above a certain degree of severity for an interest to trump a right.[11] Courts must apply 'an elaborate, sophisticated and rather strict test of justification' when a fundamental right is at stake in a specific case, while they tend to adopt 'a highly relaxed and deferential approach' when 'a "mere" individual interest is at stake.'[12] Interests that have not been elevated to the status of human rights 'will be less carefully scrutinized' than interests that have acquired such a status.[13] Treating Article 20(2) of the ICCPR as a simple additional limitation to the exercise of freedom of expression would require ascertaining the acceptable limits that may be imposed on freedom of expression when protecting the *interest* to be free from the harm of hate speech, at the adjudication level. In contrast, the recognition of that interest as a *right* would require that adjudicatory bodies, while considering hate speech cases, not just determine the limits of freedom of expression but also to strike a balance and resolve tensions between these two fundamental rights (and possibly other rights, as well, including freedoms of religion and assembly).

The right codified by Article 20(2) falls most precisely under the ambit of rights that 'maximize the *utility* of freedoms' than under the ambit of rights that 'maximize the *range* of freedoms'.[14]The right to be free from the harm of hate speech enhances and facilitates the enjoyment of the fundamental rights to individual dignity and equality that run like a red thread throughout IHRL instruments.[15]The principle affirmed in Article 5 of the ICCPR that 'no one may engage in an activity aimed at destroying the rights of others'[16] also provides the rationale for the right to be free from the harm of hate speech, given that it prohibits the abuse of freedom of expression with the aim of enhancing the rights of others.[17]

II.

Bringing together four specific internal features of the right to be free from the harm of hate speech and analysing their combined effect on the right's difficult evolution provides useful insights into the major challenges involved in this regard. Moreover, these insights prove useful beyond the confines of the international legal regulation of this issue. Indeed, any legal system that provides protection against the harm of hate speech, whether at the national, regional or

[11] Eva Brems, "Introduction," in *Conflict Between Fundamental Rights*, ed. Eva Brems (Intersentia, 2008), 2; Janneke H. Gerards, "Fundamental Rights and Other Interests: Should It Really Make a Difference?," in *Conflict Between Fundamental Rights*, ed. Eva Brems (Intersentia, 2008), 688.

[12] Gerards, "Fundamental Rights and Other Interests," 680.

[13] Ibid., 688.

[14] Feldman, *Civil Liberties and Human Rights*, 13;Feldman uses these categories to distinguish between negative rights and positive rights.Feldman, *Civil Liberties and Human Rights*, 13

[15] Nowak, *U.N. Covenant on Civil and Political Rights*, 474; Karl Josef Partsch, "Freedom of Conscience and Expression, and Political Freedoms," in *The International Bill of Rights: The Covenant on Civil and Political Rights*, ed. Louis Henkin (New York: Columbia University Press, 1981), 229; Scott J. Catlin, "Proposal for Regulating Hate Speech in the United States: Balancing Rights Under the International Covenant on Civil and Political Rights," *Notre Dame Law Review* 69 (1994 1993): 795,810; David Kretzmer, "Freedom of Speech and Racism," *Cardozo Law Review* 8 (1987 1986): 467; Andrew Altman, "Freedom of Expression and Human Rights Law: The Case of Holocaust Denial," in *Speech and Harm: Controversies over Free Speech*, ed. Ishani Maitra and Mary Kathryn McGowan (Oxford: Oxford University Press, 2012), 31.

[16] "International Covenant on Civil and Political Rights, UNGA Resolution 2200A (XXI), December 16, 1966, 21 UNGAOR Supp. (No.16) at52, UN Doc. A/6316 (1966), 999 UNTS 171, Entered into Force March 23, 1976," n.d., Article 5.

[17] Nazila Ghanea, "Expression and Hate Speech in the ICCPR: Compatible or Clashing?," *Religion and Human Rights* 5, no. 2–3 (2010): 177–178; Stephanie Farrior, "Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech," *Berkeley Journal of International Law* 14 (1996): 4–5.

international level, has to grapple with the interpretation and implementation challenges besetting these four features of the norm against hate speech.

The right's first internal feature, its 'emotional' component, manifests itself in the nature of the prohibited expressions, as well as the nature of one category of harms that justify such prohibition. More precisely, the right obliges states to make their national laws intolerant of an extreme emotion, which is *hatred*, if its advocacy incites, inter alia, *hostility*, which is an emotional harm. In this sense, the key terms 'hatred' and 'hostility' construct the meaning of the right to be free from the harm of hate speech. However, they are unrelated to concrete practice, instead connecting to invisible occurrences that concern the states of minds, attitudes and psychological states of abhorrence, detestation and enmity. The right's 'emotional' component renders the clear and objective definition of the right in the context of IHRL a difficult task. Furthermore, this 'emotional' component represents the right's relativist challenge, as emotive states are rooted in conceptions of morals – which, in turn, encompass relativity and change; they shift through time and vary from place to place. In contrast to physical or bodily harms, international norms that provide protection from emotional harms are less likely to 'resonate transnationally'.[18]

The right's second feature is the difficulty of proving the causal relationship between advocacy of hatred and its alleged harms. This causal relationship is indirect, cumulative, and mentally and emotionally mediated and thus proves difficult to precisely or empirically establish and measure. The complexities of proving the inciting nature of advocacy of hatred pose a considerable challenge to the interpretation and enforceability of the right to be free from the harm of hate speech. Furthermore, the incitement component of the right represents an additional relativist challenge to its interpretation, as the nature and strength of the causal relationship between speech and its alleged harms is contextual: the prevailing wider social environment and historical context shape both the meaning and impact of speech. Notably, international norms that do not entail 'a short and clear causal chain' connecting the source of harms to the harms themselves are less likely to cross cultures and acquire wide international agreement.[19]

The interpretation and implementation of the right to be free from the harm of hate speech require striking a very delicate balance between the speakers' and listeners' rights to equality and liberty; these tensions represent the right's third feature. These tensions are not, as they might appear prima facie and as the academic literature widely describes, between the two values of liberty and equality in the abstract. Rather, they occur between the interests of listeners and speakers in both values. The involvement of the two values (liberty and equality) in the right is multifaceted and complex. The right takes effect by restricting the speakers' freedom of expression. Yet liberty as a value is not enhanced only through the protection of a wider range of expressions; liberty can be at risk for the listeners if they are not provided protection against the harms of hate speech, as hate speech can have a 'silencing effect' on them. Enhancing the equality of listeners is the major underlying rationale of the right. However, equality as a value can be at stake for the speakers if their exercise of freedom of expression is unwarrantedly infringed upon. Thus the interpretation and implementation of the right should not be reduced to solving perceived tensions or even conflicts between equality and liberty and then giving primacy to one value over the other. Instead, the right's interpretation and implementation involve striking a difficult balance between the speakers' and listeners' interests in enjoying their liberty and equality rights.

[18] Margaret E. Keck and Kathryn Sikkink, "Transnational Advocacy Networks in International and Regional Politics," 1999, 99, http://isites.harvard.edu/fs/docs/icb.topic446176.files/Week_7/Keck_and_Sikkink_Transnational_Advocacy.pdf; See also Darren Hawkins, "Explaining Costly International Institutions: Persuasion and Enforceable Human Rights Norms," *International Studies Quarterly* 48, no. 4 (2004): 779.

[19] Keck and Sikkink, "Transnational Advocacy Networks in International and Regional Politics," 98–99.

Moreover, the right has a group-identity aspect. Both the hateful content and the harms of expressions prohibited are those targeting three specific collective identities: national origin, race, and religion. The group-identity component of the right makes the right embody another source of tension between individual and group rights, representing the right's fourth internal feature. This feature questions whether the right protects only individual members of the three identified groups or protects these groups as such from collective harms, as well. Drawing a sharp dividing line between the two categories of protection (the individualized and the collective) entails an evident difficulty given that hate speech targets people based on their group-defining characteristic(s) or identity. The group-identity component of the right also raises the dilemma of how to distinguish between the protection of groups from the collective harms of hate speech and the protection of the group-defining characteristic(s) given the blurred lines separating both kinds of protection. This dilemma has become more visible and contentious recently, especially in regards to religious hate speech, as it significantly affects the scope of prohibited expressions pursuant to the right. The fact that the right is a fusion of individual and collective elements thus poses evident challenges to its interpretation and definition.

There are interconnections between the definitional uncertainties and tensions that underlie the four internal features of the right. The 'emotional' component of the right further compounds the difficulties involved in proving and defining incitement, while these two features of the right complicate the resolution of the two sources of tensions that beset it. In addition, addressing the right's definitional uncertainties becomes more difficult given its underlying tensions. These tensions are reconciled differently, reflecting biases regarding conceptions of or approaches to rights. This, in turn, leads to varied definitions of the right's key terms, and consequently to different delineations of its exact meaning and scope. Since the regulation of hate speech involves accommodating the ideas of formal and substantive notions of equality and autonomy as well as individual and group rights, the libertarian and egalitarian conceptions of freedom of expression as well as the individualist and communitarian approaches to rights largely inform and guide the right's various modes of interpretation and implementation.

The right to be free from the harm of hate speech as codified in IHRL requires positive state action, in the form of enactment of necessary laws, (rather than just the non-interference of states) and takes effect through limiting the legitimate zone of expressions available to speakers in order to do justice to the equality principle. Furthermore, the right recognizes both emotional and tangible/physical harms of hate speech as well as group-based harms. Thus the right aligns with the egalitarian notion of freedom of expression, which recognizes a wide range of harms, including non-physical harms, that justify the restriction of hate speech by the state in order to respect the equality principle. The right also aligns with the communitarian approach to rights, which seeks to protect group identities and safeguard group rights. On the other hand, the strictly libertarian-individualist approach to rights greets the right to be free from the harm of hate speech sceptically. This approach holds that the state's prohibition of hate speech is generally an impermissible restriction upon the content of speech, except when the speech is directed toward an *individual* under circumstances in which an *immediate violent or lawless act* is likely to result. Strong biases regarding the libertarian-individualist approach to rights are irreconcilable with the approach endorsed by IHRL in regulating hate speech.

The analysis of the four internal features of the right to be free from the harm of hate speech indicates that the shift from moral denunciation to legal prohibition of hate speech at the international level entails inherent interpretive and implementation challenges. These challenges are evident when defining the content of prohibited speech, proving its causal relationship with proscribed harms, resolving tensions between the speakers' and listeners' rights to liberty and

equality, and resolving tensions between individual and group rights. The major areas of contention among states in conceiving the meaning of the right have their basis in different approaches to addressing the definitional uncertainties and tensions underlying the right's four internal features. Moreover, these definitional complexities and tensions represent the major obstacles confronting the consolidation of supra-national hate speech jurisprudence. Thus, the four internal features of the right to be free from the harm of hate speech render it challenging for the right to acquire a universal, definite and consensual meaning. These controversies and complexities are traced in the three main domains of the right's journey in IHRL.

III.

An analysis of the drafting history of the ICCPR's Article 20(2) gives rise to an examination of the emergence of the right to be free from the harm of hate speech in IHRL. The obvious links between hate propaganda and the grave human rights violations committed before and during WWII against particular groups drove the international legal regulation of hate speech. Article 20(2) was largely triggered by the tragic impact of Nazi and Fascist ideologies and policies (specifically, their hateful and racist propaganda), and the desire to prevent the resurgence of similar extreme ideologies. Moreover, the drafting of Article 20(2) took place against the ideological polarization between the East and West during the Cold War. It was promoted mostly by states belonging to or allied with the Eastern bloc and was resisted mostly by states belonging to the Western bloc.

The *travaux preparatoires* of Article 20(2) were characterized by considerable controversy: not only was the delineation of the Article's exact scope controversial, but its very inclusion in the ICCPR was similarly debatable. The states that opposed the Article's perceived it as failing to set forth a human right, but rather just as imposing unwarranted restrictions on freedom of expression. Thus, they believed that the Article did not fall under the Covenant's substantive scope. They conceived of the ICCPR as an instrument that should set forth only individual rights of a negative nature, entailing the non-interference of states. On the other hand, the states that supported the Article perceived it as indispensible for establishing a new post-war world order. They held that protection against the denial and abuses of human rights should be guaranteed for every human being and should in itself constitute a recognizable right in IHRL.

The polarized positions of the Article's supporters and opponents during the drafting phase were mainly related to the four internal features of the right to be free from the harm of hate speech. Concern about the perceived tensions between states' obligations under Article 20(2) and freedom of expression permeated the Article's drafting history. Both supporters and opponents resorted to the logic of abuse to justify their positions. The former warned against the abuse of the exercise of freedom of expression and viewed the Article as a preventive tool to provide protection to individuals and groups against discrimination. The latter, on the other hand, warned against the abuse of limitations on the fundamental freedom of expression in a manner that could lead to jeopardizing it; they viewed the Article as an abusive tool in itself.

The Article's drafting history indicates the split among states regarding whether IHRL should provide protection against the emotional harms of hate speech. Objections arose, mainly from Western states, to the use of the terms 'hatred' and 'hostility' in the Article, on the basis that the terms defy objective legal definitions and are highly susceptible to abuse by states that might impose unwarranted restrictions on freedom of expression. Western states tried to narrow the scope of harms justifying prohibiting advocacy of hatred as much as possible, and called for restricting this scope to the incitement to violence alone. They resisted specifying legal means with

which to address hate speech and regarded such specification as approaching legislating morality. Western states affirmed that the particularities of national contexts, especially histories, legal traditions and political cultures shape national approaches to addressing hate speech, and might not necessarily be legal in nature. The Article's supporters, however, insisted that IHRL provides protection against a broader range of harms resulting from advocacy of hatred. They regarded incitement to hatred or to hostility as equally dangerous as incitement to violence, on the basis that the latter usually takes root in the former, and held that the serious repercussions of the former incitements meant that these incitements deserve to be prohibited by law across states.

The multiplicity of states' understandings regarding the meaning of the Article's terms and rationale make the recourse to its *travaux preparatoires*, as a supplementary interpretive tool, not conclusively determinative of the Article's exact meaning. The voting record on the Article at the UN, where it was adopted with 50 states in favour, 18 against and 15 abstentions, indicates that the wide and deep disagreements among states on both the Article's objectives and means of achieving them, had not been resolved during negotiations. The right to be free from the harm of hate speech was thus the product of a fragile international agreement and was codified within IHRL without a shared understanding of how to address its definitional uncertainties and underlying tensions.

IV.

In light of the many definitional or conceptual ambiguities and inherent tensions that the right to be free from the harm of hate speech embodies, the development and refinement of supra-national interpretive jurisprudence is crucial in understanding the right's exact meaning and scope. Despite the fact that Article 20(2) of the ICCPR imposes a strict liability on states, the Human Rights Committee (HRCttee), as the body designated with the task of interpreting the Covenant, has not met the challenge of clarifying its exact threshold. The Committee did not develop clear criteria for defining the Article's key terms, particularly 'hatred', 'hostility', and 'incitement'. The first term describes the substantive content of expressions that fall under the Article's scope; the second describes one category of the harms that justify prohibiting these expressions; and the third is crucial in activating and enforcing the right through proving causation between expressions and their alleged harms. The HRCttee was unable to determine, or deliberately avoided determining, the exact normative content and scope of the right to be free from the harm of hate speech, and how exactly to delineate its legal contours in relation to freedom of expression.

The HRCttee was prone to regard the main legal issues that arose within the context of hate speech through the lens of freedom of expression, rather than the right to be free from the harm of hate speech. In the few individual communications on hate speech which it considered, the major question driving its legal reasoning was whether Article 19 protected the expressions involved. The HRCttee did not adopt separate legal tests to assess whether these expressions met the threshold of Article 20(2). Instead, the Committee employed the Article only as an additional or supplementary element to inform or reinforce legitimate conditions for restricting freedom of expression, particularly 'the protection of the rights and reputation of others' and 'necessity'. Overall, the HRCttee's hate speech jurisprudence has left the right to be free from the harm of hate speech ill-defined and vague (both with respect to its normative core, and as a corollary, as to what constitutes a violation).

Similar to the HRCttee, the Committee on Elimination of Racial Discrimination (CERD) and the European Court on Human Rights (ECtHR) have avoided articulating the essential definitional elements of the key terms integral to hate speech adjudication, in particular hatred and

incitement. They avoided being constrained in their reasoning by definitions which could limit their power of action in subsequent cases. The available body of supra-national hate speech jurisprudence has not provided a principled assessment of the qualification of certain expressions as advocating hatred, nor did they provide a systematic method for identifying incitement for the purpose of restricting the exercise of freedom of expression.

The HRCttee and the ECtHR have taken a cautious supra-national scrutiny approach to hate speech. They conceived of national authorities as best positioned to demarcate the boundary between free speech and hate speech, particularly regarding the qualification of the nature of prohibited expressions and the assessment of their likelihood to cause harms (their inciting nature). The analysis of causation in supra-national hate speech jurisprudence reflects a conviction that such causation is largely contingent upon national contexts. This body of jurisprudence is largely unhelpful in determining the exact elements that should characterize national laws that provide protection against the harms of hate speech. Moreover, it has thus far failed to develop clear legal tests to guide national adjudicatory bodies in interpreting and implementing hate speech laws. Supra-national hate speech jurisprudence indicates that when it comes to concretizing the right to be free from the harm of hate speech and balancing it against other rights or interests, states enjoy wide discretion. A variety of national approaches to the formulation and application of hate speech laws could be regarded as compatible with the basic guarantee of providing protection against hate speech. This was true also of the ECtHR's jurisprudence, though a higher level of homogeneity exists among the legal traditions and cultures of states parties to the ECHR.

The absence of a consolidated and comprehensive body of supra-national hate speech jurisprudence has been largely influenced by the four internal features of the norm against hate speech. Emotional harms have been recognized as grounds for restricting hate speech, however, the norm's emotional component has hampered the articulation of objective and clear definitions of the key terms involved in hate speech adjudication, particularly 'hatred' and 'hostility'. Moreover, this emotional component is largely responsible for the relaxed scrutiny approach that supra-national human rights monitoring bodies follow, as it introduces the element of legal moralism into hate speech regulation. Notably, these bodies have traditionally avoided giving clear directions to states regarding the protection of morals as legitimate grounds for restricting the exercise of freedoms, on the basis that the notion of morals should be contextually interpreted.

Furthermore, the nature of the causal relationship between advocacy of hatred and its alleged harms (which is indirect, cumulative, belief-mediated and highly contingent upon context), explains the difficulties these supra-national bodies have encountered in defining the threshold of incitement for the purposes of restricting hate speech. The complexities involved in proving incitement have also driven these bodies to adopt, to a large extent, the national authorities' assessments of how tightly the causal relationship between advocacy of hatred and its actual or potential harms must be drawn before restricting the exercise of freedom of expression. Notably, the CERD's more proactive interpretive approach, when compared to that of the HRCttee, can be understood in light of the fact that the prohibition of racist hate speech under the ICERD does not involve the complexities of proving a causal relationship between racist statements and their alleged harms. The ICERD prohibits the mere act of disseminating racist ideas, without the need for qualifiers of intention, advocacy, or incitement.

Supra-national monitoring and adjudicatory bodies have been largely unable to provide clear guidance to states on the reconciliation of the two main sources of tensions underlying the regulation of hate speech. These bodies have elaborated upon the right to be free from the harm of hate speech only as an additional legitimate limitation to freedom of expression, rather than as

an autonomous substantive right. They confronted tensions between speakers' and listeners' rights to equality and liberty that have hampered the articulation of strictly defined parameters or stringent tests to identify the threshold between freedom of expression and prohibited hate speech in supra-national jurisprudence. The complex and multifaceted involvement of the two fundamental values (liberty and equality) within the right to be free from the harm of hate speech has prompted supra-national monitoring bodies to adopt almost a purely case-based and context-based approach in adjudicating hate speech cases rather than articulating clear principles on how to balance the competing rights of speakers and listeners. These bodies have not acknowledged that they need to reach beyond a case-by-case approach.

Regarding the tensions between individual and group rights, supra-national bodies have recognized harms to groups as such, not only to their individual members, as justifying the imposition of restrictions on hate speech. The jurisprudence of these bodies aligns with group rights approach to protection against the harm of hate speech, rather than the strictly individualist approach. However, the delicate issue of distinguishing between the incitement to hatred against groups and the incitement to hatred against group-defining characteristic(s) remains largely unclear in the available jurisprudence despite the issue's rising relevance to contemporary hate speech challenges (especially in the context of religious hate speech). The incitement to hatred against groups, the incitement to hatred against the individual members of groups, and the incitement to hatred against or the denigration of group-defining characteristic represent, at the abstract level, separate analytical categories of expressive acts. However, religious, racial or national hatred actually refers to hatred against a group of persons defined by reference to (or centred around) religious belief, racial identity, or national origin, which raises doubts about whether a line could clearly be drawn between protecting groups or their individual members from incitement and protecting their identity from denigration or defamation. Moreover, in reality, there are overlaps and commonalities between these categories of expressive acts. Many groups perceive contempt or insult of their identities as amounting to incitement to hatred against them and to group defamation since they identify themselves primarily by their group identities. Also, expressions that are construed as denigrating particular group identities are frequently used as a pretext to incite hatred, discrimination or violence against groups in an indirect manner. The group-identity component of the right to be free from the harm of hate speech or in other words the fact that the right is integral to the promotion of collective goals, the prevention of communal harms, and the protection of groups' identities explains the difficulty encountered by supra-national bodies in making the distinction between the incitement to hatred against groups and the incitement to hatred against group-defining characteristic.

The analysis of the case law of the International Criminal Tribunal for Rwanda (ICTR) reveals that the difficulties confronting the consolidation of case law on incitement to genocide are less complex than those confronting the consolidation of case law on the right to be free from the harm of hate speech (given that genocide is the most public and extreme manifestation of violence and relevant case law came in response to its actual, rather than possible, occurrence). However, the difficulties confronting the two bodies of jurisprudence are not entirely distinct from each other. In both instances, the internal features of respective norms resonate (in particular the emotional component and the incitement component), leading to definitional uncertainties and complexities in proving causation between speech and its alleged harms.

These various interpretational gaps and weaknesses within the body of supra-national hate speech jurisprudence make it hard to determine when a violation has occurred in any given circumstance and as a consequence, when the right to be free from the harm of hate speech can be claimed before a court or a relevant monitoring body. Furthermore, they are not conducive to the long-term development or refinement of the right in IHRL.

V.

The analysis of recent efforts to develop the right within the UN also provides an important window through which the right and its associated dynamics can be assessed. There were persistent efforts during the last decade within the HRC that focused on developing new international standards on hate speech. These standard-setting efforts were led by Islamic states and supported by most African states; they started with a series of UN resolutions on combating defamation of religions that were successfully adopted from 1999 to 2010. These resolutions aimed to recognize the prohibition of defamation of religions as an international human right norm that should enjoy significant political and legal weight. However, given the non-binding nature of these resolutions, Islamic states regarded them as insufficient to fully address their concerns. They decided to redouble their efforts through the 2006 establishment of the UN Ad Hoc Committee on the Elaboration of Complementary Standards, which had the explicit mandate of creating new international binding standards on incitement to racial and religious hatred. The standards that Islamic states proposed aimed mainly to oblige states to prohibit by law the negative stereotyping and defamation of religions, as well as the derogatory profiling and stigmatization of both individuals and groups on the basis of religion. Islamic states first resorted to Article 20(2) to legitimize their suggested non-binding standards within resolutions on combating defamation of religions. They framed these resolutions as falling under the ambit of the right to be free from the harm of hate speech. Subsequently, they contended that the Article had normative gaps, and focused on the need to address such gaps through the development of new international binding standards complementary to Article 20(2).

While the Eastern bloc mainly initiated the codification of the right to be free from the harm of hate speech in IHRL in response to the atrocities committed before and during WWII, in recent years, Islamic states have led international advocacy efforts on the issue. These states were prompted to address the rising manifestations of Islamophobia in the West, especially in the aftermath of 9/11, with the creation of new international standards on religious hate speech. Conversely, Western states have historically been the main opponents to efforts aiming to regulate hate speech under IHRL. During the 1950s and 1960s, they resisted the codification of the right to be free from the harm of hate speech in IHRL and sought to narrow the scope of Article 20(2) of the ICCPR as much as possible. Western states have also been the main opponents to the recent efforts aiming to further expand the international norm against hate speech over the last decade.

Islamic states' suggested standards for religious hate speech sought to realign the normative scope of the right to be free from the harm of hate speech through the direct intervention with the four internal features of that right. First, the suggested standards expanded the 'emotional' component of the right by providing protection to the feelings of religious adherents and by recognizing offence and insult to them as legitimate grounds for imposing restrictions on freedom of expression. Second, they recognized an automatic causal relationship between defamation of religions or offensiveness to religious feelings on the one hand, and incitement to discrimination, hostility or violence against religious adherents on the other. The suggested standards read an implication of the right to be free from religious defamation in the original right to be free from the harm of hate speech. By implying this right, the suggested standards overlooked the difficulties in proving the causal relationship between advocacy of religious hatred and its alleged harms. Third, the Islamic states' proposed standards lowered the threshold of the legitimate exercise of freedom of expression for speakers by adding the respect of religions, their symbols and sacred personalities, and the respect of religious feelings as legitimate grounds for restricting

freedom of expression. Fourth, the suggested standards expanded the protection accorded to religious groups and their identities from hate speech in a manner inclusive of their group-defining characteristic, religion. These proposed standards' objects of protection were not simply religious adherents, but also religions themselves. This expansion of the group-identity aspect of the right to be free from the harm of hate speech reflects the Islamic states' conviction of the inseparability of incitement that targets followers of religions and that which targets the defamation of religions.

The opponents of the standard-setting agenda, on the other hand, decoupled the standards suggested by Islamic states from the scope of Article 20(2). They rejected the proposed standards' intervention with the normative content and scope of the right to be free from the harm of hate speech under IHRL, in particular rejecting the addition of the prohibition of defamation of religions as falling under the ambit of the right. The opponents of the suggested standards considered the latter an infringement on the legitimate exercise of freedom of expression. They emphasized that religions do not warrant protection under IHRL, which protects only the rights of individuals, and that those rights do not include the right to be free from insults or offences to religion. The opponents of the standard-setting agenda held that the Islamic states' proposals, if adopted, would negatively impact the international human rights system itself by distorting the IHRL's focus on individuals. Moreover, they contended that the current IHRL framework lacks any normative gaps and is sufficient to address contemporary hate speech challenges.

Historical and contemporary analyses of states' positions on how to address hate speech within IHRL help to highlight the elements of continuity and change in the areas of contention involved. A clear element of continuity has permeated these discussions: the major areas of contention among states that arose during the drafting of Article 20(2) of the ICCPR more than 60 years ago are in many ways similar to those arising during contemporary standard-setting efforts on hate speech. States' different positions in both phases stemmed from their different perceptions of how to address the definitional uncertainties and tensions underlying the four internal features of the right to be free from the harm of hate speech. However, since recent standard-setting attempts focused on the religious ground of hate speech, new controversial aspects of the right to be free from the harm of hate speech were brought to the fore. The relationship between freedom of religion and the right to be free from the harm of hate speech featured prominently in recent debates. In addition, the question of whether religious hate speech should be addressed as a contemporary manifestation of racism was at the heart of the new wave of inter-state debates on the right to be free from the harm of hate speech.

Islamic states did not simply draw an analogy between defamation of religions and incitement to religious hatred, discrimination or violence. Rather, their standards assumed an automatic causal relationship between defamation of religions and negative profiling of religious adherents on the one hand, and the infringement upon religious freedoms of believers and racism against them on the other. They highlighted that defamation of Islam has de facto infringed upon the rights of Muslims in the West to express and practise their religion. Islamic states also remarked upon the use of religious hate speech as a pretext for expressing hatred against Muslim communities in the West, on the basis of their racial or ethnic identities and not only their religious identity. The two appeals made by Islamic states to freedom of religion and the fight against racism to legitimize their standards further complicated standard-setting attempts. The opponents of the standard-setting agenda considered the Islamic states' suggested standards as a remodelling of the scope of freedom of religion and racism under IHRL. They excluded any causal relationship between defamation of religion and the violation of religious freedoms of believers and their subjection to racism. The causal relationship among these four different analytic categories (defamation of religions; incitement to religious hatred, discrimination or violence; violation of religious freedoms; and racism against believers) is not necessarily inevitable nor absolutely impossible. Instead, the

interconnection, including possible practical overlaps and links, of these four analytic categories should be empirically examined without predetermined or fixed assumptions, rather than conceived of in strictly normative or conceptual terms. No single determination can be made *in abstracto* on such complex matters.

The recent standard-setting efforts in the area of religious hate speech have created tensions at the level of multilateral human rights diplomacy between Islamic states and Western states. These tensions resulted not only from polarized positions on the need and desirability of creating new standards, but they were also exacerbated by the fact that Islamic states framed their suggested standards as a response to the deteriorating situation of Muslim minorities in the West. The statements made by Islamic states regarding the debates on standard-setting within the UN took an accusatory tone towards the West. These states aimed to exert political pressure on Western states to address the rising manifestations of Islamophobia, and to expose what they perceived to be rights deficits or protection gaps in the areas of minority rights, racism and xenophobia in the West.

After several rounds of diplomatic standoffs in the HRC during the last decade, the gaps between the supporters and opponents became too wide to bridge, and efforts aiming to develop international standards on hate speech ultimately reached an impasse. This paper has highlighted the role played by the composition and capacities of supporters and opponents in the failure of recent standard-setting efforts in the area of religious hate speech. Islamic states, supported by most African states, were the only supporters of standard-setting efforts. Islamic states failed to expand the support for their efforts from either other states or other members of the international law-making community (as NGOs or UN Special Rapporteurs or other independent human rights experts). In contrast, the opponents were much more effective than the supporters in advancing their normative agendas. In addition to Western states, the opponents included a number of Asian and Latin American states as well as active and dynamic international and national NGOs. The opponents, particularly the US, played a significant role, by investing political and diplomatic capital and exerting political pressure, in obstructing the standard-setting efforts. They first gradually reduced the level of support for these efforts by other states, and then pushed the Islamic states to freeze their initiatives.

Debates within the UN on resolutions on combating defamation of religions, as well as debates within the Ad Hoc Committee's sessions, have stimulated further intense debates on Article 20(2) of the ICCPR and have ultimately changed its dormant status within IHRL. Despite the greater visibility that the Article has acquired over the last decade, the standard-setting efforts in the area of religious hate speech exposed a polarized and confrontational reading of the Article as highly controversial within the edifice of the ICCPR. States' positions reflected the lack of a common understanding of the rationale and meaning of the right to be free from the harm of hate speech, in terms of both the nature of prohibited expressions pursuant to it and the location of the threshold for prohibiting freedom of expression. States' differences on how to address hate speech within IHRL stem from their polarized conceptions of the four internal features of the right to be free from the harm of hate speech. Such polarized perceptions of the four intrinsic qualities of the international norm against hate speech also belie the failure of the majority of states to reach an agreement regarding the further development of this norm in IHRL.

While early negotiations on the codification of the right to be free from the harm of hate speech in IHRL ended in the adoption of Article 20(2) of the ICCPR based on a fragile international agreement, the latest negotiations on the international norm against hate speech have reached an impasse.

The evolution of the international norm against hate speech within IHRL has faced significant difficulties and appears to be almost frozen in a specific frame. The right to be free from the harm of hate speech clearly presents a challenge to IHRL. States' polarized readings of the meaning of the right emerged early in the right's codification, surfacing in the period from 1947-1961 during the negotiations surrounding the drafting of Article 20(2) of the ICCPR. Furthermore, states' polarized understanding of the right has also manifested itself in the last decade during attempts to create additional international standards on hate speech in the UN. The significant controversies that characterized the drafting of Article 20(2) had not been reconciled by the time of the Covenant's adoption, while the latest standard-setting efforts to expand the right's normative content ended in a prolonged stalemate. While the history of the right's negotiations and the recent negotiations to further develop it demonstrate a considerable gap among states in their understanding of the right's rationale and scope, international jurisprudence has not helped to fill this gap. The body of hate speech jurisprudence that supra-national monitoring and adjudicatory bodies have developed has contributed very little to the elaboration of the normative content of the right, leaving a broad spectrum for states' discretionary interpretations. Thus, the exact normative core and scope of the international human right to be free from the harm of hate speech, as captured in the text of the ICCPR, continue to be contested, lacking a universal or consensual meaning. The right has not yet transformed into a common or shared understanding of its exact normative content among states, nor has it developed into specific standards within supra-national jurisprudence to guide its worldwide interpretation and effective implementation.

VI.

The analysis provided in this paper on the emergence of the right to be free from the harm of hate speech in IHRL, its interpretive supra-national jurisprudence, and recent efforts to expand its normative content, has demonstrated that this right has presented a challenge to IHRL. Despite the right's codification more than sixty years ago via Article 20(2) of the ICCPR, its normative content remains largely unsettled and underdeveloped. The four internal features of the right have played an influential role in shaping its difficult evolution within IHRL. The translation of the claim to be, or interest in being, free from the harm of hate speech into an international human right carries significant interpretive and implementation challenges. While Article 20(2) established a strong international legal norm obliging, and not just authorizing, states to prohibit by law the expressive acts described therein, four intrinsic qualities of that norm subsequently contributed to its lack of a coherent and universal meaning. These are: the emotional component; the incitement component; tensions between speakers' and listeners' rights to liberty and equality; and the group-identity component.

While these four internal features of the right to be free from the harm of hate speech have influenced the difficult path the right has taken so far in its evolution within IHRL, they also shape the right's prospects for further development and inform its expansive potential. Drawing upon the analysis of the right's four internal features, this paper submits that any possible efforts to create new international standards against hate speech that aim to expand the right's normative core by adding specificity to the content and effects of the proscribed expressions pursuant to it will confront multiple difficulties.

States' polarized positions on how to address the complex definitional challenges and tensions underlying the four internal features of the international norm against hate speech constitute a significant obstacle to its expansive capacity. Both during the drafting of Article 20(2) and during the recent standard-setting attempts in the area of religious hate speech, it became clear that efforts to reach a wide agreement among states on how to regulate hate speech under IHRL would

ultimately reach an impasse. Moreover, the weaknesses and interpretational gaps that characterize supra-national hate speech jurisprudence indicate that any possible standard-setting efforts would be confronted by a dearth of clear and well-established sets of principles that could be built upon or developed in the form of new international standards.

The existence of a minimum level of uniformity across a critical mass of states in implementing and interpreting the *original* obligatory standard established by Article 20(2) is essential for the development of *new* international standards against hate speech. However, national approaches to hate speech regulation show significant variations across states that apply different criteria, in both legislative patterns and judicial practices, to define the threshold between free speech and hate speech. The OHCHR recently affirmed this, after its comprehensive assessment of the state of implementation of the prohibition of incitement to national, racial, or religious discrimination, hostility, or violence, as outlined in Article 20(2) of the ICCPR, in a variety of countries around the world.[20] NGO ARTICLE 19 after undertaking a similar assessment of the laws and jurisprudence on hate speech across countries in different regions of the world characterized them as a 'patchwork' where it found significant variations 'in how prohibition and threshold of incitement is approached and defined in laws and regulations, and in how these concepts are applied.'[21]

Drawing upon this paper's analysis of the internal features of the right to be free from the harm of hate speech, it is clear that both the 'emotional' and incitement components of the right make the right's interpretation and implementation largely contextual. Furthermore, legal traditions across states resolve the two sources of tensions that the right embodies between speakers' and listeners' rights to equality and liberty, as well as individual and group rights, differently. These differences emanate from the legal traditions' biases to either egalitarian or libertarian notions of freedom of expression, and their different levels of commitment to the advancement of group rights. Consequently, this leads to varied delineations of the meaning and scope of the right to be free from the harm of hate speech across countries that are shaped by, in addition to the various legal traditions, cumulative practical experiences and particularized political, cultural, and historical national settings. More specifically, the legislative patterns and judicial practices of the resolution of the hate speech problem are predicated upon different conceptions of: the content of prohibited expressions; the scope of recognized harms of hate speech; the extension of the right's protection to groups and to group-defining characteristics; the range of groups protected; and the standards of causality between advocacy of hatred and its alleged harms. The prohibition of advocacy of hatred that constitutes clear and unambiguous incitement to immediate violence or illegal acts is the aspect of the right to be free from the harm of hate speech that enjoys transnational resonance, since it easily crosses cultural and ideological boundaries. However, the legal regulation of advocacy of hatred that falls short of incitement to violence but does create a

[20] "Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence," October 5, 2012, para. 11, http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf.

[21] Henry Maina, "The Prohibition of Incitement to Hatred in Africa: Comparative Review and Proposal for a Threshold" (presented at the OHCHR Expert workshops on the prohibition of incitement to national, racial or religious hatred, Workshop for Africa, Nairobi, 2011), http://www.ohchr.org/EN/Issues/FreedomOpinion/Articles19-20/Pages/ExpertsPapersNairobi.aspx; Anges Callamard, "Towards an Interpretation of Article 20 of the ICCPR: Thresholds for the Prohibition of Incitement to Hatred" (presented at the OHCHR Expert workshops on the prohibition of incitement to national, racial or religious hatred, Vienna, 2011), http://www.ohchr.org/Documents/Issues/Expression/ICCPR/Vienna/CRP7Callamard.pdf; Sim Kok Eng Amy, "Preventing Hatred or Silencing Voices: Making the Case for a Rigorous Threshold for the Incitement to National, Racial or Religious Hatred" (presented at the OHCHR Expert workshops on the prohibition of incitement to national, racial or religious hatred, Workshop for Asia Pacific, Bangkok, 2011), http://www.ohchr.org/Documents/Issues/Expression/ICCPR/Bangkok/AmySim.pdf; Paula Martins, "Freedom of Expression and Equality: The Prohibition of Incitement to Hatred in Latin America," 2011, http://www.ohchr.org/Documents/Issues/Expression/ICCPR/Santiago/PaulaMartins.pdf.

social climate conducive to hostility and discrimination does not enjoy the same universal resonance. Instead, it is subject to different states' approaches.

There is no doubt that the right to be free from the harm of hate speech has so far resisted substantive evolution and refinement. However, states are still in need of further guidance in the resolution of boundary disputes in hate speech regulation, especially in light of the current globalizing context of hate speech which poses a new set of human rights challenges. After its recent examination of the state of implementation of Article 20(2) of the ICCPR worldwide, the OHCHR has characterized legislations in the area of hate speech as often 'excessively narrow or vague' and has characterized related jurisprudence as 'scarce and ad hoc'.[22] Furthermore, the OHCHR has noted the existence of a 'dichotomy' between the lack of prosecution of actual incitement cases and 'persecution of minorities under the guise of domestic incitement laws'.[23] After making a similar assessment, the NGO ARTICLE 19 characterized laws related to incitement as 'inconsistent and vague in their application'. It furthermore characterized the available jurisprudence at national levels as 'vague, ad hoc and possibly lacking in conceptual discipline or rigour.'[24] As Frank La Rue, the UN Special Rapporteur on freedom of expression, rightly notes, the different national regulatory responses to hate speech are 'symptomatic of the unclear [international] normative environment surrounding the issue.'[25] As this paper has illustrated, the scant guidance supra-national jurisprudence on hate speech has provided on the interpretation of the international norm against hate speech has compounded the definitional uncertainties of this norm. The ambiguities that surround the interpretation and implementation of the international norm against hate speech open the door to excessive prohibitions, inconsistent implementation and restrictive interpretations, without effective scrutiny from supra-national monitoring bodies.[26]

In order to address the resistance of the international norm against hate speech to substantive evolution, efforts aiming to further develop this norm, in response to contemporary hate speech challenges, can be directed toward approaches that place less emphasis on legal and textual development and more on providing practical guidance to states about implementing the norm. Given the formulation of the right to be free from the harm of hate speech, any new international binding standards against hate speech would oblige states to add new elements to their national offences, incurring liability on speakers engaging in hate speech. However, the possible margin of agreement among states on the *legal* measures for addressing hate speech is very narrow. Moreover, the interaction of the various rights and interests involved in hate speech regulation presents scenarios and complications that are so rich and diverse that a ready-made formula to resolve them universally is hardly feasible. Indeed, appeals to context frequently arose during the right's drafting history and in recent UN debates. Supra-national human rights monitoring bodies also largely endorsed these appeals, which seemed to function as valid reasons for favouring one normative interpretation of the right over another.

Despite the limited expansive potential of the international norm against hate speech, the greater visibility it has acquired within the lexicon of IHRL has created the momentum to develop a set of principles providing practical guidance to states on how to draw the line between freedom of

[22] "Rabat Plan of Action," para. 11,15.
[23] Ibid., para. 11.
[24] Maina, "The Prohibition of Incitement to Hatred in Africa"; Callamard, "Towards an Interpretation of Article 20 of the ICCPR"; Amy, "Preventing Hatred"; Martins, "Freedom of Expression and Equality: The Prohibition of Incitement to Hatred in Latin America."
[25] *Report of Frank La Rue the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, September 7, 2012, para. 3, http://daccess-dds-ny.un.org/doc/UNDOC/GEN/N12/501/25/PDF/N1250125.pdf?OpenElement.
[26] See: ARTICLE 19, "Prohibiting Incitement to Discrimination, Hostility or Violence," 2012, http://www.article19.org/data/files/medialibrary/3572/12-12-01-PO-incitement-WEB.pdf.

expression and prohibited incitement. The elaboration of guiding principles on the implementation of the right to be free from the harm of hate speech, that transcend contextual considerations or that are sufficiently open to accommodate highly relevant contextual variables, present one solution for bypassing political impasses on the right's substantive evolution. This would add much-needed discipline, consistency and rigor to the methodologies that courts employ to reach their rulings on hate speech cases, and can also guide legislators in drafting relevant hate speech offences in their national legislations.

The '*Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence*',[27] which the OHCHR issued in 2012, can be considered the basis for consensual guiding principles, if built upon. The plan proposed a six-part threshold test as a framework for the implementation of the right to be free from the harm of hate speech or determining expressions that warrant prohibition under Article 20(2) of the ICCPR. A number of UN Special Rapporteurs, human rights experts and NGOs have endorsed this threshold test,[28] which is composed of the following six elements. First, the social and political *context* prevalent at the time the expression was made and disseminated, in terms of the existence of conflicts or tensions within society, among groups; present or historic acts of violence or discrimination targeting members of particular groups; or frequent negative stereotyping of particular groups. Second, the *speaker,* in terms of his position or status and his influence or authority over the audience. Third, the *intent*, which could be indicated by analysing the language the speaker uses, as well as the scale and repetition of expressions. Fourth, the *content* of expression, in terms of the degree to which the expression was provocative and direct; the severity of the harm advocated; and the form, style, and nature of the arguments used. Fifth, the *extent* of speech, in terms of its reach, public nature, magnitude, frequency and the medium of dissemination. Sixth, the *likelihood* of the occurrence of the harm in terms of the existence of its degree of risk and reasonable probability.

These six criteria, designed to assist states with the establishment of boundaries between freedom of expression and prohibited incitement under Article 20(2), represent a true opportunity for a systematic engagement with some of the challenging aspects of the right to be free from the harm of hate speech. They can serve as a focal point to bring international players away from political posturing and towards convergence on a focused agenda, helping granting the right a streamlined or harmonious interpretation and implementation. In contrast to the creation of international substantive standards outlining the content of prohibited expressions and their undesired harms pursuant to the right, the procedural development of the right, or outlining the basic guarantees for its realization, is not overloaded with navigational challenges. This procedural approach could substitute a strong normative content to the right that is universal or consensual.

The articulation of *guiding* principles for determining whether certain expressions are prohibited pursuant to the international norm against hate speech, that leave enough room for a contextual assessment, could be the way forward to further develop the norm and enhance its realization since such principles could garner wide endorsement from states. The specification of procedural aspects of the legal interpretation of the international norm against hate speech helps to maintain and protect a minimum universal core of this norm, instead of the mere acceptance of the norm's fragmentation into separate historically-based, culturally-defined, politically-shaped, and country-specific approaches.

[27] "Rabat Plan of Action."
[28] Ibid., para. 22.

This paper has shed light on the significant difficulties facing the development, strengthening and expansion of the international norm against hate speech. It has provided an explanatory framework through which to understand the origins and the causes of core challenges facing the evolution of the international norm against hate speech. It contends that four internal features of this norm (the emotional component, the incitement component, tensions between speakers' and listeners' rights to liberty and equality, and the group-identity component) have a strong and direct influence on understanding the difficult path the norm has taken in its evolution within IHRL.

This analysis has illustrated the importance of examining the role that the internal features of international human rights could have in influencing their normative evolution. It provides a framework from which further reflection and scholarship on the right to be free from the harm of hate speech, and by extension on the evolution of other human rights, can emerge. The examination of the push and pull factors for normative evolution in IHRL have not yet featured as a major research question within international relations and international law disciplines. Additional empirical studies on this area of research, which focus not only on cases that witnessed successful normative evolution but also on cases in which such evolution seems to face serious obstacles or resistance, is needed to contribute to its theoretical development. As the human rights challenges that have global dimensions, and thus global impact, grow, questions about the normative expansion of the existing international human rights framework also acquire significant policy relevance. Explaining the internal and external dynamics of normative evolution in the context of IHRL is crucial to the identification of its conditions and consequently to the formulation of strategies for different stakeholders that can shape and influence such an evolution.

UNIVERSITY OF CAMBRIDGE

Centre of Governance and Human Rights

Citation: Elbahtimy, M., (Jan. 2014) 'The Right to be Free from the Harm of Hate Speech in International Human Rights Law', *CGHR Working Paper 7*, Cambridge: University of Cambridge Centre of Governance and Human Rights