UN COMMITTEE ON THE ELIMINATION OF RACIAL DISCRIMINATION

81st Session, 6-31 August 2012

Day of Thematic Discussion

"Racist Hate Speech"

Palais des Nations, Geneva, 28 August 2012

The troubled relationship between free speech and racist hate speech:

the ambiguous roles of the media and internet

Dr. Tarlach McGonagle[*]

Introduction[1]

The Committee on the Elimination of Racial Discrimination (hereafter, 'CERD' or 'the Committee') is to be commended for the sense of vision and purpose that it has shown in organising a Thematic Discussion on "Racist Hate Speech". This represents a determined attempt to engage frontally with a particular, virulent strain of hate speech.

Racist hate speech shares many of the characteristics of hate speech *simpliciter*. It is, for instance, a vexed notion. It means many things to many people and inevitably some of those meanings are contestable and contested. It is also a recalcitrant notion. Neither the term nor the types of expression it denotes are likely to go away in the foreseeable future. Furthermore, racist hate speech is, politically-speaking, a highly-charged notion. There is notable divergence across constitutional, legislative, political and socio-cultural approaches to racist hate speech at the national level. Such divergence can prove frictional in particular circumstances (e.g. high-profile controversies with a transfrontier character) or contexts (e.g.

---

the drafting of the International Convention for the Elimination of All Forms of Racial Discrimination (hereafter, 'ICERD')).

The Thematic Discussion seeks to elucidate the term, "racist hate speech", and to explore in depth and in detail the specific relevance of the term in the context of ICERD. As such, the exercise is as necessary as it is timely. Its necessity stems from two observations. First, ICERD is widely – and correctly - perceived as an outlier among other international human rights treaties that contain provisions governing the relationship between freedom of expression and hate speech, insofar as Article 4, ICERD, creates more far-reaching obligations for States parties than comparable provisions in other treaties. Article 4, ICERD, requires States to render several types of expression *punishable* by law, whereas Article 20 of the International Covenant on Civil and Political Rights (hereafter, 'ICCPR') requires that a narrower range of types of expression be *prohibited* by law. Moreover, the complex and somewhat confusing provisions in ICERD concerning racist hate speech, and the (overly?) strict manner in which they have traditionally been interpreted by CERD,[2] both underscore the need for insightful clarification of key phrases.

An exploration of the relationships between racist hate speech, the media and internet will feature centrally in this paper. However, at the very outset, it is important to also recall and insist upon the importance of the crucial relationships that underpin and inform those relationships: *viz.*, the relationships between freedom of expression and racist hate speech and between freedom of expression and all kinds of media, including internet-based media. The former relationship will be dealt with immediately in Section I, because it is of fundamental importance, whereas the latter two relationships will be teased out at relevant stages in Sections II and III, respectively.

I. Freedom of Expression and (Racist) Hate Speech: Relational and Definitional Tensions

---

[2] For detailed analysis, see: Patrick Thornberry, "Forms of Hate Speech and the Convention on the Elimination of all Forms of Racial Discrimination", 5 *Religion and Human Rights* (2010), pp. 97-117; Patrick Thornberry, "Confronting Racial Discrimination: A CERD Perspective", 5 *Human Rights Law Review* (No. 2, 2005), pp. 239-269; Tarlach McGonagle, *Minority Rights, Freedom of Expression and of the Media: Dynamics and Dilemmas*, *op. cit.*, pp. 280-290.

"Hate speech" is less a term of art than a term of convenience. It cannot properly be considered a term of art because it has not (yet) been defined in a watertight or authoritative way, either in international human rights law or in relevant scholarship. It is more fitting to describe 'hate speech' as a term of convenience because of its ability to refer to a broad category of expression in a shorthand way. Its appeal is obvious. In contradistinction to precise but cumbersome legal jargon, "hate speech" is a term that is readily understood as referring to objectionable, harmful and (very often) illegal types of expression motivated by one kind of animus or another. This point can be illustrated nicely in the context of ICERD. It is much more convenient to simply refer to 'racist hate speech' than to itemize all of the different types of expression envisaged by Article 4, ICERD:

> States Parties condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in article 5 of this Convention, inter alia:
>
> (a) Shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof;
>
> (b) Shall declare illegal and prohibit organizations, and also organized and all other propaganda activities, which promote and incite racial discrimination, and shall recognize participation in such organizations or activities as an offence punishable by law;
>
> (c) Shall not permit public authorities or public institutions, national or local, to promote or incite racial discrimination.

But we need to be careful whenever we use the term "racist hate speech". Legal precision and nuance should not be traded in for rhetorical convenience. The term's straightforward appearance belies the great complexity of its actual scope. "Hate speech" can actually be taken as referring to a whole spectrum of (extremely) negative discourse stretching from hatred and incitement to hatred; to abuse, vilification, insults and offensive words and epithets; and arguably also to extreme examples of prejudice and bias.[3] Robert Post has posited that a certain threshold of intensity must be reached before a particular expression can

---

[3] James B. Jacobs and Kimberly Potter, *Hate Crimes: Criminal Law and Identity Politics* (New York, Oxford University Press, 1998), p. 11.

be qualified as hate speech.[4] He points to the *Oxford English Dictionary* entry for "hate": "an emotion of extreme dislike or aversion; detestation, abhorrence, hatred".[5] For Post, the threshold or definitional prerequisite is the qualification, "extreme", because ordinary "intolerance and dislike are necessary human emotions which no legal order could pretend to abolish".[6]

From a legal perspective, the aforementioned spectrum of negative discourse stretches from types of expression that are not entitled to protection under international human rights law (e.g. incitement to hatred), through types of expression that may or may not be entitled to protection, depending on a number of "contextual variables"[7] (e.g. extremely offensive expression), to types of expression that presumptively would be entitled to protection, despite their objectionable character (e.g. negative stereotyping of minorities). The right to freedom of expression necessarily covers expression that may "offend, shock or disturb" certain groups in society (which is not the same thing as a *right* to offend[8]).[9] Democracy is not without its rough edges and tough talk is part of the cut and thrust of public debate and discourse.

Intuitively, there can be no objection to Bhikhu Parekh's condemnation of hate speech as "objectionable for both intrinsic and instrumental reasons, for what it is or manifests and for what it does".[10] However, the shift from the moral condemnation of hate speech to its legal regulation (or prohibition) inevitably requires definitional clarity and circumspection. As Jean-François Gaudreault-DesBiens has observed:

> The multiple forms of anti-egalitarian expression that exist are neither equally harmful nor
> performative; we must not, therefore, lose sight of the link between the norm that the state is

---

[4] Robert Post, "Hate Speech", in Ivan Hare and James Weinstein, Eds., *Extreme Speech and Democracy* (New York, Oxford University Press, 2009), p. 123-138, at 123.
[5] *Ibid*.
[6] *Ibid*.
[7] Michel Rosenfeld, "Hate Speech in Constitutional Jurisprudence: A Comparative Analysis", 24 *Cardozo Law Review* (2003), 1523-1567, at 1565.
[8] See further, Onora O'Neill, "A Right to Offend?", *The Guardian*, 13 February 2006; Francesca Klug, "Freedom of Expression Must Include a Licence to Offend", 1(3) *Religion and Human Rights* (2006), 225-227.
[9] *Handyside v. the United Kingdom*, Judgment of the European Court of Human Rights of 7 December 1976, Series A, No. 24, para. 49.
[10] Bhikhu Parekh, "Is There a Case for Banning Hate Speech?", in Michael Herz & Peter Molnar, Eds., *Content and Context: Rethinking Regulation and Remedies for Hate Speech*, *op. cit*., pp. 37-56, at 44.

drafting and the broader public policies involved when identifying [*sic*] the specific forms of anti-egalitarian expressions to discourage.[11]

Further differentiation between forms of hate speech has usefully been provided by Richard Delgado and Jean Stefancic, as follows: "direct (sometimes called 'specific') or indirect; veiled or overt; single or repeated; backed by power, authority, or threat, or not".[12] These types of differentiation are crucial when attempting to gauge the impact of racist hate speech on its targets/victims.

Once the differentiation inherent in the term hate speech and its significance have been recognized and understood, meaningful examination of the rationales for regulating hate speech can commence.

The purpose of regulating hate speech is to prevent interference with other rights and to prevent the occasioning of certain harms. In the first place, hate speech can interfere with other human rights or "operative public" values:[13] dignity, non-discrimination and equality, (effective) participation in public life (including public discourse[14]), freedom of expression, association, religion, etc. Second, the prevention of particular harms suffered by individual victims should also be considered: psychological harm, damage to self-esteem, inhibited self-fulfilment, etc.[15]

All in all, the range of harms to be prevented is varied and complex. The challenge is therefore to identify "which criteria allow us to distinguish between harms that justify

---

[11] Jean-François Gaudreault-DesBiens, "From Sisyphus's Dilemma to Sisyphus's Duty? A Meditation on the Regulation of Hate Propaganda in Relation to Hate Crimes and Genocide", 46 *McGill Law Journal* (2000), 121-139, at 133. See also in this connection, Bhikhu Parekh, "Is There a Case for Banning Hate Speech?", *op. cit.*, at 54-55.

[12] Richard Delgado and Jean Stefancic, "Four Observations about Hate Speech", 44 *Wake Forest Law Review* (2009), 353-370, at 361. See also: Richard Delgado and Jean Stefancic, *Understanding Words That Wound* (Boulder, Colorado, Westview Press, 2004), pp. 11-12.

[13] Operative public values are those values "that a society cherishes as part of its collective identity and in terms of which it regulates the relations between its members", and which "constitute the moral structure of its public life and give it coherence and stability": Bhikhu Parekh, *Rethinking Multiculturalism: Cultural Diversity and Political Theory* (2nd Edition) (New York, Palgrave Macmillan, 2006), p. 363.

[14] For a thorough analysis of this topic, see: Robert C. Post, "Racist Speech, Democracy, and the First Amendment", 32 *William and Mary Law Review* (1991), 267-327.

[15] See generally, Mari J. Matsuda, Charles R. Lawrence III, Richard Delgado and Kimberlè Williams Crenshaw, Eds., *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Boulder/San Francisco/Oxford, Westview Press, 1993).

restrictions and those that do not".[16] Those criteria should then guide relevant regulatory approaches to hate speech. Whereas some types of racist hate speech – the most egregious forms - may be best dealt with by regulatory (including criminal) measures, others are more suitably dealt with by educational, cultural, informational and other non-regulatory (and necessarily non-criminal) measures. Insofar as a regulatory framework is necessary to counter racist hate speech, that framework should be holistic, in recognition of the fact that racist hate speech covers a range of different types of expression. But it is not enough for that regulatory framework to be holistic: the approaches it sets out must also be differentiated. The "horses for courses" principle applies.

Implications for ICERD

The foregoing observations have a number of implications for the Committee's interpretation of ICERD in its monitoring and related activities. First, it is important to catch hold of the notion, racist hate speech, and draw it in before it escapes our grasp entirely. The term needs to be aligned, or rather, re-aligned with the precise contours of the Convention's provisions dealing with relevant issues. It is important to stress that this concerns more than one provision: thus, not only Article 4 (*juncto* Article 5), but also Article 7, ICERD. Following that re-alignment, a very clear assessment will have to be made of which types of expression are covered by relevant provisions of the Convention, and which types are not. The specific types of expression that are covered by the Convention will then have to be addressed in a much more targeted way than would ever be possible under the vague and expansive notion of "racist hate speech". The challenges of re-alignment and focused engagement are considerable, but CERD is not alone in facing them; similar gauntlets are waiting to be picked up by other UN treaty bodies such as the Human Rights Committee, which only dealt in summary fashion with the relationship between Articles 19 and 20, ICCPR, in its General Comment No. 34 on the right to freedom of expression.[17]

It has already been suggested, above, that there are important differences between intrinsic and instrumental rationales for regulating (racist) hate speech. Whereas the *intrinsic* reasons for devising regulatory and other measures to combat (racist) hate speech remain largely

---

[16] David Kretzmer, "Freedom of Speech and Racism", 8 *Cardozo Law Review* (1987), 445-513, at 478.
[17] United Nations Human Rights Committee, General Comment No. 34 – Article 19: Freedoms of opinion and expression, Doc. No. CCPR/C/GC/34, 12 September 2011, paras. 50-52.

unaffected by advances in information and communications technology, the *instrumental* reasons for doing so are altered considerably. The following sections seek to identify a number of reasons why this is so in respect of the media and internet. The role the media can play in promoting or preventing racist hate speech is of central importance.[18] With the inexorable rise of Internet, the media are currently undergoing rapid and sweeping changes, with the result that their potential and actual roles in spreading or countering racist hate speech need urgent reassessment. Technological and related new dynamics have also changed the regulatory and policy agendas for protecting and promoting freedom of expression and combating racist hate speech.

II. Media

Before scrutinizing the relationship between the media and racist hate speech, it is useful to recall the prior relationship between freedom of expression and the media. This is a complicated relationship that is defined by mutual dependencies. On the one hand, free expression is a prerequisite for the media to be able to operate freely. On the other hand, (levels of) access to media structures and processes and access to media output can determine whether the right to freedom of expression – in particular its component rights to seek, receive and impart information and ideas of all kinds – is exercised effectively by individuals and groups in practice. This can be explained by the media's reach and impact, which will be elaborated on below.

Little reminder is needed of how the media can contribute to the dissemination of racist hate speech: concrete examples abound, both in contemporary times and historically. The abuse of mass media for totalitarian propaganda in the run-in to, and during, the Second World War has led to a legacy of (pre-)caution regarding the right to freedom of expression. The enduring apprehension of the double-edged nature of freedom of expression (and the media) prompted the drafters of the ICCPR to refer to freedom of expression as being simultaneously a "precious heritage" and a "dangerous instrument".[19] This apprehension explains why, for

---

[18] Hans Christian Krüger, "Use of the Media to Promote and Infringe Human Rights", in Kathleen E. Mahoney and Paul Mahoney, Eds., *Human Rights in the Twenty-First Century* (The Netherlands, Kluwer, 1993), pp. 743-756.

[19] Marc Bossuyt, *Guide to the "Travaux Préparatoires" of the International Covenant on Civil and Political Rights*, (Dordrecht, etc., Martinus Nijhoff Publishers, 1987), p. 386.

instance, "special duties and responsibilities" are considered to be an integral part of the right to freedom of expression and govern its exercise (see, for example, Article 19(3), ICCPR).

Although it is understandable that the relationship between the media and racist hate speech is often cast in negative terms, the reality of how the media engage with racist hate speech is somewhat more ambiguous and more complex than the mere provision of channels for the dissemination of hateful messages. The media can, of course, also be used to counter racist hate speech. The ambiguity of the media's roles flows from the double-edged nature of freedom of expression and the complexity of their roles is aptly demonstrated by the *Jersild v. Denmark* case, which led to a seminal judgment by the European Court of Human Rights.[20]

The case involved the conviction of Jens Olaf Jersild, a Danish journalist, for aiding and abetting in the dissemination of racist statements in a televised interview he had conducted. The statements in question were uttered by members of an extreme right-wing group known as the "Greenjackets" and the journalist was convicted largely because he had failed to explicitly contradict, or distance himself from, the racist and xenophobic statements of the interviewees. The European Court of Human Rights held that Jersild's conviction was not "necessary in a democratic society" and that it therefore violated his rights under Article 10 of the European Convention on Human Rights. This conclusion rested largely on considerations of context in (news) reporting and the importance of journalistic autonomy for the functioning of democracy. The Court held that the journalist's right to freedom of expression had been infringed, *inter alia*, because it was not for the courts to determine which journalistic techniques (e.g. "the methods of objective and balanced reporting") should be used.[21]

The *Jersild* principles provide an excellent point of departure for one of the main arguments advanced in this paper, *viz.*, that it is very important to move beyond an enduring apprehension of freedom of expression and the media towards an appreciation of how freedom of expression and the media can (be used to) counter racist hate speech. In order to appreciate the potential role that the media can play in that regard, it is first necessary to understand some relevant media dynamics.

---

[20] *Jersild v. Denmark*, Judgment of the European Court of Human Rights of 23 September 1994, Series A, No. 298.
[21] *Ibid.*, para. 31.

The influence and power wielded by the media in society derives largely from their reach and impact. The societal presence of the media is virtually ubiquitous and they have properly been described by Roger Silverstone as constituting "an essential dimension of contemporary experience".[22] Ed Baker has called the media "the central *institution* of a democratic public sphere".[23] It is easy to understand the premises on which these observations have been made. In practice, information and ideas are circulated, and public debate is conducted, primarily via the media. Thus, the media serve as channels (and amplifiers) for information and ideas and as discursive forums. By virtue of their reach, speed, influence and impact, more often than not, the media are the most effective means of seeking, receiving and imparting information and ideas of all kinds. The effective exercise of the right to freedom of expression is therefore often contingent on effective access to the media and an effective ability to receive content via the media.

Roger Silverstone has explained with great verve the intricate interplay of factors that assure the media's influence over individual and collective opinion-making processes:

> It's all about power, of course. In the end. The power the media have to set an agenda. The power they have to destroy one. The power they have to influence and change the political process. The power to enable, to inform. The power to deceive. The power to shift the balance of power: between state and citizen; between country and country; between producer and consumer. And the power that they are denied: by the state, by the market, by the resistant or resisting audience, citizen, consumer. It is all about ownership and control: the who and the what and the how of it. And it is about the drip, drip, drip of ideology as well as the shock of the luminous event. It is about the media's power to create and sustain meanings; to persuade, endorse and reinforce. The power to undermine and reassure. It is about reach. And it is about representation: the ability to present, reveal, explain; and also the ability to grant access and participation. It is about the power to listen and the power to speak and be heard. The power to prompt and guide reflection and reflexivity. The power to tell tales and articulate memories.[24]

---

[22] Roger Silverstone, *Why Study the Media?* (London/Thousand Oaks, CA/New Delhi, SAGE Publications, 1999), p. 1. See also, Peter Dahlgren, *Television and the Public Sphere: Citizenship, Democracy and the Media* (London, SAGE Publications, 1995), p. 155.
[23] (emphasis per original). C. Edwin Baker, "Viewpoint Diversity and Media Ownership", 60 *Federal Communications Law Journal* (No. 3, 2009), 651-671, at 654.
[24] Roger Silverstone, *Why Study the Media?*, *op. cit.*, p. 143.

When analysing this intricate interplay of factors, it is instructive to also probe into the theories of media effects, reception analysis and communicative power.[25] This makes for vastly more sophisticated analysis than bland, incantatory assumptions about the impact and reach of the media.

Widespread under-appreciation of the complex nature and impact of the media means that the positive role that the media can play in countering hate speech is often under-appreciated as well. It is frequently argued that more speech or counter-speech[26] is the best antidote to racist hate speech. This argument is sometimes disputed in respect of individual, face-to-face situations, but the focus here will be on the broader societal ramifications of counter-speech. The media, thanks to their ability to disseminate and amplify information and ideas, and to provide shared platforms for deliberative interaction, are crucial facilitators of counter-speech. To properly gauge their importance in this respect, it is necessary to understand counter-speech not only as reactive or responsive to concrete instances of racist hate speech, but as a pro-active or pre-emptive force.

Insofar as minority cultures, identities and languages are often marginalized, disadvantaged or discriminated against, corrective measures are required in order to ensure their preservation, transmission and development. Here the media can play an ameliorating role as important vectors of culture, identity and language. This role is very important for *intra-group* communication, but the media's forum-providing role is also very important for *inter-group* communication. Minority cultures, identities and languages undergo "public validation" when they feature in (mainstream) media.[27] Minorities are empowered through their use of mainstream media insofar as they are able to write and share their own narratives, which often challenge dominant societal perceptions and biases. Access to, and participation in, mainstream media can therefore foster inter-group deliberation, and (ideally) also inter-

---

[25] See generally, Denis McQuail, *McQuail's Mass Communication Theory* (6th Edition) (Los Angeles, etc., SAGE Publications, 2010) and in respect of communicative power, Manuel Castells, *Communication Power* (New York, Oxford University Press, 2009).

[26] For a rich theoretical exposition of counter speech and the question of its suitability as a remedy for hate speech, see: Katharine Gelber, *Speaking Back: The Free Speech Versus Hate Speech Debate* (Amsterdam/Philadelphia, John Benjamins Publishing Company, 2002), and Katharine Gelber, "Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia)", in Michael Herz & Peter Molnar, Eds., *Content and Context: Rethinking Regulation and Remedies for Hate Speech*, *op. cit.*, pp. 198-216.

[27] Stephen Harold Riggins, "The Promise and Limits of Ethnic Minority Media", in Stephen Harold Riggins, Ed., *Ethnic Minority Media: An International Perspective* (USA, SAGE Publications, 1992), pp. 276-287, at 283.

group understanding and tolerance. The dynamics of dialogical interaction can help to pre-empt racist hate speech by seeking to use deliberation to prevent the escalation of differences or tensions that commonly give rise to racist hate speech.

Firm faith in the pre-emptive power of counter-speech and inter-group dialogue has shaped the Council of Europe's comprehensive set of strategies against racist hate speech. The Council's approach is also characterized by a differentiated understanding of racist hate speech and is best described as a "horses for courses" approach.[28] As such, it involves targeted strategies for combating distinct types of expression, alternating—depending on the situation—between strategies that are restrictive of certain types of expression and strategies that seek to promote other types of expression or expressive opportunities, especially via the media. Within the Council of Europe, engagement with relevant issues by the Advisory Committee on the Framework Convention for the Protection of National Minorities and the European Commission against Racism and Intolerance has been particularly detailed. The engagement of those two bodies, in the context of their monitoring activities, has facilitated the identification of relevant best practices for the media.

Implications for ICERD

The foregoing observations and reflections are of clear relevance for CERD as it tries to anticipate how relevant issues will play out in the future and how to steer its continued engagement with those issues. Building on the conclusions to Section I of this paper, the challenge of ensuring that the media play an effective role in combating racist hate speech calls primarily for promotional measures, not prescriptive ones. As such, this is a golden opportunity for Article 7, ICERD – with its emphases on educational, cultural and informational strategies – to truly come into its own. Article 7 reads, in full:

> States Parties undertake to adopt immediate and effective measures, particularly in the fields of teaching, education, culture and information, with a view to combating prejudices which lead to racial discrimination and to promoting understanding, tolerance and friendship among nations and racial or ethnical groups, as well as to propagating the purposes and principles of the Charter of the United Nations, the Universal Declaration of Human Rights, the United Nations Declaration on the Elimination of All Forms of Racial Discrimination, and this Convention.

---

[28] See further, Tarlach McGonagle, "A Survey and Critical Analysis of Council of Europe Strategies for Countering 'Hate Speech'", *op. cit.*

Article 7 could help to harness the potential of the media – arising from their channelling/amplifying and forum-providing capacities - for fostering intercultural communication and understanding. In order to do so, Article 7's high-minded objectives would need to be translated into firm measures that are meaningful and effective in practice. Guidance and inspiration for such a translation exercise could usefully be sought in the range of best practices distilled from various standard-setting and monitoring activities of the Council of Europe, as discussed above. This could help to change the perception of Article 7 as ICERD's "neglected pillar".[29]

III. Internet

If there ever was a time when the media could accurately have been described as a unified, homogenous entity, it is long gone. Today, heterogeneity in the media is the order of the day. The unitary appearance and catch-all character of the term 'media' masks a more complex range of different media types, the most readily distinguishable of which include public service, commercial, community, local, transnational, etc. While broad-brush and summary, this typology suffices at least to indicate that different media types have different objectives, target audiences and levels of geographical reach.

Moreover, in recent years, due mainly to the advent and relentless growth of the internet, the media have been undergoing profound changes; they are generally becoming increasingly instantaneous, international and interactive.[30] In tandem, ideas, information and content of all kinds are generally becoming more abundant, accessible and amplified to wider sections of society. As a result of these changes, the current media offering is more plentiful and varied than it has been at any point in history. These developments have prompted observations that internet content is "as diverse as human thought".[31] There is now a greater range of media at our disposal than ever before, offering wider and more diversified functionalities/capabilities and greater differentiation in types of access, participation and output.

---

[29] Stephanie Farrior, "The Neglected Pillar: The "Teaching Tolerance" Provision of the International Convention on the Elimination of All Forms of Racial Discrimination", 5 *ILSA J. Int'l & Comp. L.* 291 (1999), 291-299.

[30] See generally, Karol Jakubowicz, *A New Notion of Media? Media and Media-Like Content and Activities on New Communications Services* (Strasbourg, Council of Europe, 2009).

[31] *ACLU v. Reno*, 929 F. Supp. 824 (11 June 1996), 842, para. 74.

These advances in information and communications technologies can clearly have far-reaching consequences for how racist hate speech is disseminated and processed. The internet holds unprecedented potential for multi-directional communicative activity: unlike traditional media, it entails relatively low entry barriers. Whereas in the past it was necessary to negotiate one's way through the institutionalized media in order to get one's message to the masses, this is no longer the case. There is reduced dependence on traditional points of mediation and anyone can, in principle, set up a website or communicate via social media. Messages therefore can - and do - spread like wildfire across the globe. Often, all that is needed for a message to "go viral" is a combination of strategy and happenstance. While there are no guarantees that an individual's message will actually reach vast international audiences, the capacity to communicate on such a scale clearly does now exist for an ever-expanding section of the population.

The accessibility and effectiveness of the internet as a medium of communication is largely due to the ease, speed and versatility with which expression can be disseminated online. It is increasingly[32] being used for spreading racist hate speech in different ways and contexts, including:

- dissemination of propaganda, other types of (mis-)information and hate spam;
- exchange of information and ideas, e.g. via social media networks, discussion groups, listservs and communities of interest;
- attracting inadvertent users by "usurping domain names" and "using misleading meta-tags";[33]
- organizational purposes such as the coordination of activities, planning of events, training, recruitment drives;
- commercial ends such as fund-raising, the sale of publications, videos, memorabilia and paraphernalia;

---

[32] Empirical studies of the growing online presence of hate speech are surveyed *inter alia* in: Barbara Perry and Patrik Olsson, "Cyberhate: The Globalization of Hate", 18(2) *Information & Communications Technology Law* (2009), 185-199; James Banks, "Regulating Hate Speech Online", 24(3) *International Review of Law, Computers & Technology* (2010), 233-239; Yaman Akdeniz, *Racism on the Internet* (Strasbourg, Council of Europe Publishing, 2009). Organizations active in the monitoring of online hate speech include the Simon Wiesenthal Center (<www.wiesenthal.com>) and the Anti-Defamation League (<www.adl.org>).
[33] Elizabeth Phillips Marsh, "Purveyors of Hate on the Internet: Are We Ready for Hate Spam?", 17 *Georgia State University Law Review* (2000)*,* 379-407, at 391.

- pursuit of various offences against the person and other criminal or invasive behaviour, e.g. the targeting of (potential) victims, cyber-bullying, cyber-stalking.

From a regulatory perspective, new technological possibilities and how they are exploited in practice present a number of complicating factors. The first cluster of factors can be grouped around liability and jurisdictional issues and the second cluster comprises factors affecting victims of hate speech.

Liability and Jurisdictional Issues

Owing to the virtual, globalized and decentralized features of the architecture of the internet, online hate-mongers enjoy a high degree of mobility. These technological features allow such hate-mongers to offer content via Internet Service Providers (ISPs) based in a jurisdiction of their choice. This is clearly relevant for the struggle against online hate speech because national laws can vary quite considerably in the extent to which they tolerate hate speech. In the United States, for instance, the free speech tradition cultivated by a robustly-worded First Amendment, has resulted in a very strong presumption of constitutional protection for hate speech.[34]

It is common practice for racist hate websites to be hosted in jurisdictions that are considered to be favourable to, or tolerant of, hate speech. The practice of strategically choosing favourable jurisdictions in which to host a site is sometimes called forum-shopping. It leads to regulatory circumvention and attempts to evade legal liability for hateful content. Similarly, it is also common practice for racist hate websites that have been either blocked or banned in one jurisdiction to subsequently relocate to another, more favourable, jurisdiction.[35] The ease with which racist hate websites can relocate in this manner means

---

[34] The First Amendment to the US Constitution reads: "Congress shall make no law [...] abridging the freedom of speech, or of the press [...]". For overviews and analysis of relevant First Amendment jurisprudence, see: Anthony Lewis, *Freedom for the Thought That We Hate: A Biography of the First Amendment* (New York, Basic Books, 2007); James Weinstein, "An Overview of American Free Speech Doctrine and its Application to Extreme Speech", in Ivan Hare and James Weinstein, Eds., *Extreme Speech and Democracy*, *op. cit.*, pp. 81-91; and Steven J. Heyman, "Hate Speech, Public Discourse, and the First Amendment", in *ibid.*, pp. 158-181.

[35] An example is the website of Holocaust denier, Ernst Zündel. For background and analysis, see: Yaman Akdeniz, "Stocktaking on Efforts to Combat Racism on the Internet", Background Paper for the High Level Seminar of the Intergovernmental Working Group on the Effective Implementation of the Durban Declaration and Programme of Action, Fourth Session (16-27 January 2006), United Nations Commission on Human Rights Doc. No. E/CN.4/2006/WG.21/BP.1, 9 January 2006, 16-18.

that blocking or banning cannot be regarded as effective measures against, or remedies for, such sites.

Whereas different jurisdictions have different laws governing hate speech, different ISPs within a given jurisdiction may well have different policies on hate speech too. Some ISPs' terms of service contracts and notice-and-take-down policies are more detailed and/or stringent as regards hate speech than others.

Determining legal liability for racist hate speech online is not only complicated from a jurisdictional perspective. Technological considerations also cause a number of complications in practice. Potentially, a multiplicity of different actors could be involved in the creation and dissemination of racist content: creating or sourcing it; publishing it; developing it; hosting it or otherwise facilitating its dissemination, accessibility or retrievability. Liability could attach to each of the implicated actors in different ways, depending on the nature of the communication; the scope and details of relevant national laws, and other contextual variables. These questions are enormous and enormously complicated and as such, it is beyond the scope of this paper to offer a fuller exploration of their intricacies.

Another problematic aspect of liability and jurisdictional issues concerns prosecution vagaries. Episodic or unsuccessful prosecutions have little deterrence value as they give rise to claims that relevant laws are paper tigers or toothless bulldogs. Conversely, however, overzealous prosecution can have serious chilling effects on freedom of expression and public debate. The tendency of "hate speech" laws to be formulated in terms that are over-broad, has long been a concern of some civil rights organizations and academic commentators. Vague and overbroad statutes can, for instance, be abused in order to stifle hard-edged political criticism.

Victims' Perspectives

As noted in Section I, above, it is important to be aware of the differentiation inherent in the term "racist hate speech" when assessing the harms it occasions and when calibrating relevant regulatory and other responses. Harms caused by racist hate speech and the resultant suffering of victims can be intensified by circumstances that are born out of technological capabilities or consequences. Thus, the relative ease of maintaining anonymity in an online

environment can contribute to an exacerbation of the emotional or psychological harm inflicted on victims of racist hate speech. For instance, when the true identities of those responsible for cyber-bullying, or hateful messages disseminated by mobile phone texts or via social networks, are cloaked in anonymity or pseudonymity, the very suspicion that those persons may live nearby the victim, or frequent the same social, educational or professional circles, is likely to compound his/her distress. Similarly, when messages of racist hate are circulated via social networking sites, the actual amplification of those messages, coupled with a perception that their dissemination is uncontrollable, can also increase victims' distress levels.

The potential permanency of content made available online is also a relevant consideration when quantifying the nature and extent of the harms caused by racist hate speech. Online manifestations of racist hate speech are generally more refractory than their traditional, offline equivalents. This has given rise to the term, "cyber-cesspools",[36] which conjures up the image of putrid, stagnant pools that pose a danger to public health. The durability of online content, facilitated in the first place by an absence of storage limitations, is also assured by hyperlinking and online searchability. Content remains traceable and largely retrievable after its original dissemination to an unprecedented extent when that dissemination takes place online. This means that there is a danger that victims of racist hate speech will continuously, or at least repeatedly, be confronted by the same instances of racist hate speech after their original articulation. Leading critical race theorists have argued cogently that the "incessant and compounding" aspects of racist hate speech exacerbate its impact.[37] If multi- or cross-posting or extensive hyperlinking has taken place, the removal of particular material from a particular online source cannot guarantee the unavailability of the same material elsewhere, thus strengthening its "incessant and compounding" aspects.

Responses and Remedies

Again, due to the inherent differentiation in racist hate speech, a variety of regulatory measures, including criminal law provisions, are typically employed to combat online hate speech. However, besides regulatory measures, a range of alternative and additional

---

[36] Brian Leiter, "Cleaning Cyber-Cesspools: Google and Free Speech", in Saul Levmore and Martha C. Nussbaum, Eds., *The Offensive Internet: Speech, Privacy, and Reputation* (Cambridge, Massachusetts and London, Harvard University Press, 2010), pp. 155-173.

[37] Richard Delgado and Jean Stefancic, "Four Observations about Hate Speech", *op. cit*., at 367-368.

approaches can also be suited to the specifics of different types of racist hate speech in an online setting. Such measures include: "the option of doing nothing, social norms, self-regulation, co-regulation, and technical means, information, education and awareness campaigns".[38] They can offer a number of advantages; for instance, they can be "less costly, more flexible and quicker to adopt than prescriptive government legislation".[39] However, neither regulatory measures nor any of the other measures discussed are without their shortcomings.

Against a background of scepticism regarding the effectiveness of non-regulatory measures for combating online racist hate speech, it is useful to flag a few examples of good practices. Various instances of fruitful collaboration between civil society interest groups and individual ISPs or content providers in combating racist hate speech have been documented.[40] Typically, such collaborative initiatives involve the former seeking to promote greater social responsibility on the part of the latter, by promoting (awareness of) reporting mechanisms for illegal material offered (by third parties) on their services. Another example of good practice is the International Association of Internet Hotlines (INHOPE),[41] which provides an extremely important service by enabling members of the public to anonymously report online content that they suspect to be illegal (especially child sexual abuse material, but also illegal types of hate speech). INHOPE hotlines "ensure that the matter is investigated and if found to be illegal the information will be passed to the relevant Law Enforcement Agency and in many cases the Internet Service Provider hosting the content".[42] Notwithstanding these examples of good practices for combating online racist hate speech, there remain general problems of transparency, consistency and enforceability concerning self- and co-regulatory mechanisms and processes governing ISPs.

---

[38] Yaman Akdeniz, "Introduction", in *Legal Instruments for Combating Racism on the Internet*, Report prepared and updated by the Swiss Institute of Comparative Law, Lausanne (Strasbourg, Council of Europe Publishing, 2009), pp. 7-37, at 28; Yaman Akdeniz, "Governance of Hate Speech on the Internet in Europe", in Christian Möller and Arnaud Amouroux, Eds., *Governing the Internet: Freedom and Regulation in the OSCE Region* (Vienna, OSCE Representative on Freedom of the Media, 2007), pp. 91-117, at 108. For a more detailed discussion of these measures, see: Yaman Akdeniz, "Governing Racist Content on the Internet: National and International Responses", 56 *University of New Brunswick Law Journal* (2007), 103-161, at 140-158; Yaman Akdeniz, *Racism on the Internet*, *op. cit.*, pp. 111-134.

[39] *Ibid*.

[40] For examples and analysis, see: Jessica S. Henry, "Beyond Free Speech: Novel Approaches to Hate on the Internet in the United States", 18(2) *Information and Communications Technology Law* (2009), 235-251. For an exploration of what more might be done in this regard, see: Brian Leiter, "Cleaning Cyber-Cesspools: Google and Free Speech", *op. cit.*, at 169-172.

[41] Official website available at <www.inhope.org>.

[42] *Ibid*.

The suitability of counter-speech, or more specifically, intergroup communication, strategies for combating racist hate speech was stressed in Section II of this paper. The effectiveness of counter-speech as a remedy for racist hate speech in an online environment is perhaps less self-evident than it is in the physical world. As a result of vastly enhanced communicative opportunities enabling individuals to connect with multitudes of other individuals, it seems plausible that changing patterns of individual, intra-group and intergroup communication will become discernible between off- and online variants. These developments prompt a need for fresh reflection on the effectiveness of continued normative reliance on the empowering and identity-sustaining properties of freedom of expression in an online environment.

It may, at first glance, seem paradoxical to suggest that counter-speech is likely to be less effective in an environment of informational abundance. Yet that abundance includes an abundance of racist hate speech, the pervasiveness and permanence of which is assured by the internet's archiving, hyperlinking and searching capabilities.[43] Whether overall informational abundance will drown out the abundance of racist hate speech, or dilute its impact, is too broad a question to answer *in abstracto*.

Another relevant consideration is that enhanced individual selection and filtering capacities allow individuals to choose (or "pull") their own content instead of having particular content "pushed" towards them by general intermediaries, as the institutionalized media have traditionally done. These capacities increase the ability of individuals to avoid exposure to particular types of content. The broader consequence of this is that they also reduce the chances of conflicting opinions meeting each other head-on in an online environment.[44] Such individual selection and filtering capacities can affect communicative practices at a societal level in different ways. Growing reliance on these capacities can lead to the creation of a multitude of "public sphericules" instead of a unified public sphere[45] and lead to the proliferation of communities of interest in which ideological insulation and intensification take place. The online forums in which particular types of information and especially

---

[43] Elizabeth Phillips Marsh, "Purveyors of Hate on the Internet: Are We Ready for Hate Spam?", *op. cit*., at 391.
[44] For a general discussion of selection and filtering issues concerning the Internet, see: Jonathan Zittrain, "A History of Online Gatekeeping", 19(2) *Harvard Journal of Law and Technology* (2006), 253-298.
[45] See further: Todd Gitlin, "Public Sphere or Public Sphericules?", in Tamar Liebes and James Curran, Eds., *Media, Ritual, Identity* (London, Routledge, 1998), pp. 168-175.

viewpoints are reinforced by their amplification have been described as "online echo chambers".[46]

As a result of these informational and communicative trends, the likelihood of intergroup engagement and interaction in cyberspace cannot simply be assumed; its potential is significantly reduced, compared with the offline, real-world context. Granted, "alternative (mini-) spheres"[47] can prove vitally important for intragroup communication, for purveyors of hate and minority groups alike. Some empirical research even suggests that deliberation in online echo chambers does not necessarily/always lead to more entrenched/extreme positions and that intra-group deliberation can benefit inter-group deliberation.[48] Nevertheless, in order for more speech or counter-speech strategies to have any prospect of fostering tolerance, there must be, as a minimum, communicative intent and actual communicative contact.

The failure of internet-based expression to achieve linkage to "the general public domain"[49] could lead to communication being predominantly spatial and insufficiently social. Online hate speech *has* real-life consequences, as explained above,[50] so it is crucial for online counter-speech to also realize its potential for offline effects. The promotion of targeted educational, media literacy and journalistic training initiatives could all help to create such linkage in practice. As David Heyd has astutely pointed out, "education to toleration requires the development of open-mindedness, critical scepticism, the power of deliberation, and the willingness to change one's attitude".[51] This resonates very loudly with the view that democratic society cannot exist unless it is underpinned by "pluralism, tolerance and broadmindedness".[52] By circulating information and ideas throughout society and by providing forums for dialogical interaction, the media can certainly serve these goals,

---

[46] See further, Cass R. Sunstein, *Republic.com 2.0* (Princeton NJ, Princeton University Press, 2007).
[47] Donald R. Browne, *Ethnic Minorities, Electronic Media and the Public Sphere: A Comparative Approach* (Cresskill, New Jersey, Hampton Press, Inc., 2005), p. 11. For a broad discussion of relevant issues, see: John Downing and Charles Husband, *Representing 'Race': Racisms, Ethnicities and Media* (London, SAGE Publications, 2005), esp. Chapter 9, "The Multi-Ethnic Public Sphere and Differentiated Citizenship".
[48] See, for example: Cass R. Sunstein, "Ideological Amplification", 14(2) *Constellations* (2007), 273-279; Cass R. Sunstein, *Why Groups Go to Extremes* (Washington, D.C., The AEI Press, 2008).
[49] Myria Georgiou and Eugenia Siapera, "Introduction: Revising multiculturalism", 2(3) *International Journal of Media and Cultural Politics* (2006), 243-247, at 246.
[50] See further, Alexander Tsesis, "Hate in Cyberspace: Regulating Hate Speech on the Internet", 38 *San Diego Law Review* (2001), 817-874, at 836 *et seq*.
[51] David Heyd, "Education to Tolerance: Some Philosophical Obstacles and their Resolution", in Catriona McKinnon and Dario Castiglione, Eds., *The Culture of Toleration in Diverse Societies: Reasonable Tolerance* (Manchester and New York, Manchester University Press, 2003), pp. 196-207, at 204.
[52] This is a central creed of the European Court of Human Rights. See, for example, its judgment, *Handyside v. the United Kingdom*, *op. cit*., para. 49.

including in an online environment. However, relevant policies and strategies will have to be carefully tailored to the specificities of the online context.

Implications for ICERD

As already noted, successive waves of technological developments, especially and most recently, the advent of the internet, have profoundly altered informational and communicative realities throughout the world. Those changes were not only *unforeseen* when ICERD and the ICCPR were being drafted, they were probably also *unforeseeable*. Consequently, prior understandings of the scope of the right to freedom of expression require urgent updating, adaptation and expansion in order to take account of, and accurately reflect, the complexities of the new communicative dispensation and their impact on the realization of the right to freedom of expression and other rights in practice. These observations have two concrete corollaries:

(i)   It is important that the ongoing normative evolution of the right to freedom of expression informs CERD's interpretation of the "due regard" clause in Article 4, ICERD, specifically as applied to Article 5(d)(viii), ICERD. This requires *outward-looking and forward-looking stances*; a willingness to keep abreast of interpretive and standard-setting developments elsewhere (e.g. in other international treaty bodies such as the Human Rights Committee) and to anticipate future directions in the normative evolution of the right.

(ii)   It will be important for CERD to *move beyond its own experiential horizons* when addressing media- and internet-related issues in the future. The Committee will have to re-assess and expand its prior experience of such issues in light of the increasing complexity and continued ambiguity in the roles played by the media and internet in spreading and countering racist hate speech.

Conclusions

The Thematic Discussion offers CERD a valuable moment for focused reflection on how it should engage with racist hate speech in its activities in the future. The farther shore of a new General Recommendation is reachable from here, but the journey presents considerable navigational challenges.

Differentiation needs to be recognized and appreciated on a number of levels: in the types of expression that could be categorized as racist hate speech, their intensity, the nature of the harms they occasion and the nature of the regulatory and other responses that those harms elicit. The protection of human rights and public values also calls for differentiated strategies for combating racist hate speech, including regulatory, educational, cultural and informational measures. All of these levels of differentiation remain relevant in an online context.

The term, racist hate speech, needs to be mapped onto the contours of relevant treaty provisions, in order to ascertain which provisions are applicable, when and how. Like other treaties, ICERD is not static; it is growing with the times. CERD's interpretation of the "due regard" clause in Article 4 needs to grow organically too, but in a way that is cognizant of evolving understandings of the right to freedom of expression in international human rights law. This underscores the need for CERD to adopt *outward-looking and forward-looking stances* in respect of relevant global and regional trends and developments.

The mere prospect of developing a new General Recommendation on racist hate speech invites CERD to mount a critical examination of its own positioning on relevant issues in the past. It calls for the identification of consistency (and divergence) in existing interpretive patterns and the consolidation of previous experience, with a view to more systematic engagement with relevant issues in the future. That engagement will have to reconcile formal (but imperfect) treaty provisions and objectives with a sensitivity to contextual variables, such as the complex operational dynamics of the media and internet. This is necessary in order to effectively combat racist hate speech, while safeguarding the right to freedom of expression. In particular, the empowering potential of freedom of expression will have to be nurtured in order to offset its harmful potential. Structured opportunities for intergroup communication via the media and online are crucial mechanisms for advancing that goal in practice.

All in all, the challenges of systematic engagement with racist hate speech within the parameters of ICERD will lead to an expansion of the Committee's current *experiential horizons*. That, in itself, will make the farther shore all the more reachable.