# Hate Speech: Can the International Rules be Reconciled?

## Introduction

Hate speech is one of a small number of rights issues in relation to which international law prescribes two very different sets of rules, one found in Article 4 of the *International Convention on the Elimination of all Forms of Racial Discrimination* (CERD)[1] and the other in Articles 19 and 20 of the *International Covenant on Civil and Political Rights* (ICCPR).[2] When writing about hate speech, most authors tend to gloss over the differences between these two legal regimes, or try to reconcile them, often in every general terms.

It is submitted that these two sets of rules are fundamentally incompatible. If correct, this is a serious problem for human rights proponents. There is very little chance of either treaty being amended or replaced at this point, and inconsistencies critically undermine the strength and seriousness of international human rights law, which already suffers from weak enforcement mechanisms and often flagrant breaches of the rules. Inconsistencies are even more problematical in relation to hate speech, where there are fundamental disagreements even among established democracies as to its proper scope.

These factors no doubt rank important among the reasons why few authors have tackled this difficult problem front on. But they do not justify lax or self-serving analysis of these important human rights rules, or in any way avoid the need to confront problems directly.

## The CERD Regime

CERD was the first international treaty to deal directly with the issue of hate speech and its provisions on hate speech are also by far the most far-reaching. Article 4 provides:

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

[1] General Assembly Resolution 2106A(XX), 21 December 1965, entered into force 4 January 1969.
[2] General Assembly Resolution 2200A (XXI), 16 December 1966, entered into force 3 January 1976.

> 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.

The specific obligations provided for in Article 4(a) have been analysed differently by different authors but it is probably useful to distinguish six categories of activity that States Parties are bound to declare offences punishable by law:[3]

1. dissemination of ideas based on racial superiority;
2. dissemination of ideas based on racial hatred;
3. incitement to racial discrimination;
4. acts of racially motivated violence;
5. incitement to acts of racially motivated violence; and
6. the provision of assistance, including of a financial nature, to racist activities.[4]

Four of the six activities stipulated in Article 4(a) are hate speech provisions, namely (1) to (3) and (5). Articles 4(a)(1) and (2) prohibit the mere dissemination of ideas based on racial superiority or hatred. These provisions require neither intent nor incitement to any result, whether that be an act (for example, of discrimination or violence) or simply a state of mind (i.e. hatred). Articles 4(a)(3) and (5), on the other hand, are restricted to acts of incitement, respectively to discrimination and violence, although they do not appear to include a requirement of intent.

In terms of subject matter, Article 4 refers variously (and inconsistently) to race, colour and ethnic origin, but Article 1 of CERD clearly defines 'racial discrimination' as including distinctions based on "race, colour, descent, or national or ethnic origin" so it may reasonably be assumed that other references in the Convention to 'race' extend to these categories as well.

CERD, by virtue of its specific focus on racial discrimination, does not guarantee the right to freedom of expression. However, it does require that measures taken pursuant to Article 4 have due regard for the principles set out in the UDHR – which include equality, non-discrimination and freedom of expression – and in Article 5 of CERD, which provides for equality before the law in the enjoyment of a large number of rights, including freedom of expression.[5]

We can therefore assume that these other rights may be understood as exerting some sort of moderating effect on the measures to be taken to give effect to Article

---

[3] In its General Comment No. 15 of 23 March 1993, the CERD Committee refers to four categories to be banned under Article 4. See para. 3. Lerner, N., *The U.N. Convention on the Elimination of all Forms of Racial Discrimination* (Alphen aan den Rijn, The Netherlands: Sijthoff & Noordhoff, 1980), on the other hand, identifies five categories. See p. 49.

[4] See, for example, Mahalic, D. and Mahalic, J. "The Limitation Provisions of the International Convention on the Elimination of All Forms of Racial Discrimination" (1987) 9 *Human Rights Quarterly* 74, p. 93.

[5] Article 5(d)(viii).

4. However, it would be contrary to basic principles of treaty interpretation to suggest that these references to other rights, including the right to freedom of expression, could be intended to affect the plain language understanding of Article 4; accommodation of other rights at that level should have been done at the time of drafting of these articles.

**The ICCPR Regime**

The main rule on hate speech in the ICCPR is found at Article 20(2), which provides:

> Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.

In terms of substance, Article 20(2) prohibits advocacy that constitutes incitement. It is fairly clear that 'advocacy' in this context incorporates an intent requirement. In *Faurisson v. France*, the UN Human Rights Committee was called upon to assess a conviction under a law which prohibited any contestation of the existence of the category of crimes against humanity defined in the Nuremburg Charter.[6] In a concurring opinion in that case, Evatt, Kretzmer and Klein expressed concern that the law pursuant to which the author of the complaint was convicted did "not link liability to the intent of the author".[7] However, based on the facts, the Committee decided that Faurisson had clearly been motivated by a desire to promote racism and, as a result, the conviction in that particular case was legitimate.[8]

This conclusion was expressed even more directly in *Jersild v. Denmark*, a case before the European Court of Human Rights (ECHR).[9] Formally, the ECHR was assessing whether or not the conviction of a journalist for disseminating racist statements made by others was a breach of his right to freedom of expression as guaranteed under the European Convention on Human Rights, and therefore not applying Article 20(2) of the ICCPR. However, the Court was at pains to take into account the wider international legal framework and so the reasoning is relevant here. In finding that there had been a breach of Jersild's right to freedom of expression, the Court relied heavily on its finding that Jersild's purpose or intent was not to promote racism but, on the contrary, to expose and analyse it, stating:

> [A]n important factor in the Court's evaluation will be whether the item in question, when considered as a whole, appeared from an objective point of view to have had as its purpose the propagation of racist views and ideas.[10]

The question of what constitutes incitement is extremely complex and controversial; the UN High Commissioner for Human Rights, for example, has expressed concern at the fact that the term lacks a clear definition in international

---

[6] 8 November 1986, Communication No. 550/1993.
[7] *Ibid.*, para. 9.
[8] *Ibid.*, para. 10. See also the concurring decision of Lallah, paras. 6 and 9.
[9] 22 August 1994, Application No. 15890/89.
[10] *Ibid.*, para. 31.

law.[11] It is beyond the scope of this paper to analyse this notion in detail. It is sufficient for present purposes to note that while it is not the same thing as causation; causation, if demonstrated, proves incitement (i.e. if a causal link is established between certain statements and a proscribed result, such as hostility, the statements incited that result).

However, causation is normally very difficult to prove, and courts sometimes accept general results, such as an increase in evidence of hatred among those to whom the statements were directed, as proof of incitement. Thus, in the case of *Ross v. Canada*, a teacher was removed from the classroom for his anti-Semitic/Holocaust denial publications. The Supreme Court of Canada noted the evidence that a 'poisoned environment' had been created within the relevant school board and held that "it is possible to 'reasonably anticipate' the causal relationship" between that environment and the author's publications.[12] The HRC held that this satisfied the necessity part of the test for restrictions on freedom of expression and that, as a result, there was no breach of this right.[13]

In other cases, international bodies are prepared to assume that statements which are sufficiently offensively racist, taking into account the context, meet the standard of incitement. Thus, in the *Faurisson* case, the HRC noted that the impugned statements, "were of a nature as to raise or strengthen anti-Semitic feelings".[14]

In terms of its subject matter, Article 20(2) is relatively straightforward. It prohibits statements that advocate "national, racial or religious hatred" and which incite to "discrimination, hostility or violence". In terms of the proscribed results, it is clear that hostility is a state of mind, rather than a specific action. In this regard, Article 20(2) is different from Article 13(5) of the *American Convention on Human Rights* (ACHR),[15] which only prohibits incitement to "lawless violence or to any other similar illegal action" (i.e. acts).

The rule expressed in Article 20(2) must be read in conjunction with the ICCPR's guarantees of freedom of expression and, in particular the regime for restrictions on freedom of expression, of which Article 20(2) is a special case, found at Article 19(3), which reads as follows:

> The exercise of the rights provided for in paragraph 2 of this article [which guarantees freedom of expression] carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
> (a) For respect of the rights or reputations of others;

---

[11] Study of the United Nations High Commissioner for Human Rights compiling existing legislations and jurisprudence concerning defamation of and contempt for religions, UN Doc. A/HRC/9/25, 5 September 2008, para. 24.
[12] *Ross v. New Brunswick School District No. 15*, [1996] 1 S.C.R. 825, para. 101.
[13] *Ross v. Canada*, 18 October 2000, Communication No. 736/1997, para. 11.6.
[14] Note 6, para. 9.6.
[15] Adopted 22 November 1969, entered into force 18 July 1978.

(b) For the protection of national security or of public order (*ordre public*), or of public health or morals.

The UN Human Rights Committee (HRC) has specifically stated that Article 20(2) is compatible with Article 19.[16] As a result, any law seeking to implement the provisions of Article 20(2) must not overstep the permissible scope of restrictions on freedom of expression allowed by Article 19(3). On this, the HRC has stated: "As such, a limitation that is justified on the basis of article 20 must also comply with article 19, paragraph 3."[17] This also finds support in the case law. In the case of *Ross v. Canada*, for example, the HRC specifically held that a restriction on racist expression had to be justified by reference to the test set out in Article 19(3) of the ICCPR.[18]

An interesting question is whether Article 19(3) would permit restrictions on hate speech beyond the scope of what Article 20(2) requires. Theoretically, this is possible: what States are required to ban to ensure equality is not necessarily the same as what they are permitted to ban to serve this goal without breaching the right to freedom of expression. At the same time, the drafting history of Article 20(2) suggests that there was little scope for extending its provisions while still respecting Article 19. Proposals to restrict Article 20(2) to incitement to violence were rejected, but so were proposals to extend it, for example to include 'racial exclusiveness', on the basis of concern about free speech.[19] This suggests that the obligations of Article 20(2) are either identical or extremely close to the permissions of 19(3).

*Faurisson v. France*, a case in which the HRC was called upon to assess a conviction under a law which prohibited any contestation of the existence of the category of crimes against humanity defined in the Nuremburg Charter, sheds important light on this issue. Evatt, Kretzmer and Klein, in a concurring opinion, stated:

> [T]here may be circumstances in which the right of a person to be free from incitement to discrimination on grounds of race, religion or national origins cannot be fully protected by a narrow, explicit law on incitement that falls precisely within the boundaries of article 20, paragraph 2. This is the case where ... statements that do not meet the strict legal criteria of incitement can be shown to constitute part of a pattern of incitement against a given racial, religious or national group, or where those interested in spreading hostility and hatred adopt sophisticated forms of speech that are not punishable under the law against racial incitement, even though their effect may be as pernicious as explicit incitement, if not more so.[20]

This can be understood as a call for an extremely narrow and precise interpretation of incitement in Article 20(2), alongside a recognition that there may be special

---

[16] General Comment 34: Freedoms of Opinion and Expression, 21 July 2011.
[17] *Ibid.*, para. 50.
[18] 18 October 2000, Communication No. 736/1997, para. 11.1.
[19] Bossuyt, note **Error! Bookmark not defined.**, pp. 404-405, 408.
[20] 8 November 1986, Communication No. 550/1993, para. 4.

cases where statements which do not fall within the scope of this very narrow interpretation may still legitimately be prohibited because, in context and alongside other statements, they in fact constitute a pattern of incitement. Alternately, their point could be understood as a comment on the proper interpretation of incitement, rather than going outside of the boundaries of Article 20(2), *per se*. Either way, the analysis confirms that Article 19(3) and Article 20(2) are legally contiguous or very nearly so.

**A Comparison**

The comparison of the provisions of CERD and the ICCPR on hate speech here is restricted to pointing out areas where they fundamentally disagree. The analysis takes as its starting point the idea set out in the previous paragraph that Article 20(2) describes not only the minimum standards for restricting hate speech under the ICCPR but also the maximum (or at least near maximum) scope of such restrictions. It is submitted that a principled analysis of what sorts of restrictions should be permitted in accordance with Article 19(3) of the ICCPR leads to the same conclusion.

It is fairly apparent from the analysis of the respective CERD and ICCPR hate speech regimes that there are certain fundamental incompatibilities. There are some differences in the subject matter of these regimes, inasmuch as CERD applies to racial discrimination, defined relatively broadly, while the ICCPR applies to national, racial and religious discrimination, but these do not create incompatibilities, in particular inasmuch as there is nothing in either CERD or the ICCPR to suggest that the categories to which hate speech rules apply must necessarily be limited to the ones found in either Article 4 of CERD or Article 20(2) of the ICCPR (possible other categories might include gender and/or sexual orientation).

Far more problematical, however, are Articles 4(a)(1) and (2) of CERD, which prohibit the mere dissemination of racist ideas. The centrality of intent (or advocacy) to the Article 20(2) of the ICCPR regime has already been noted. Furthermore, there are serious problems from the perspective of Article 19(3) of the ICCPR with any rule which prohibits the dissemination of hate speech in the absence of intent. The same is true of incitement, which is also central to Article 20(2) and would be required for a restriction to qualify as necessary under Article 19(3). It is submitted that the lack of intent and incitement requirements under Articles 4(a)(1) and (2) of CERD render these provisions incompatible with Articles 19 and 20 of the ICCPR.

Article 4(a)(1) of CERD controversially calls for the banning of all ideas based on racial superiority, which raises the thorny issue of when apparently positive statements about groups may constitute hate speech. At some point, and in certain contexts, advocacy of superiority may be tantamount to advocacy of inferiority and even of hatred. If so, it stands to be treated in the same way, and pursuant to the same provisions, as negative forms of hate speech. But, by terms, the superiority provision in Article 4(a)(1) is not limited in this way and its language would not

easily support this interpretation. As a result, Article 4(a) of CERD actually prohibits a wider range of positive than negative statements.

According to Mahalic and Mahalic, the UN Committee on the Elimination of Racial Discrimination has never been adverse to the dissemination of ideas which stress (positive) cultural, as opposed to racial, differences.[21] This distinction, however, would appear to be largely specious – or at least highly subjective – in many contexts where race and culture are largely synonymous. Furthermore, CERD defines race to include national and ethnic origin, which significantly overlap with culture, further obscuring the distinction.

The International Criminal Tribunal for Rwanda (ICTR), in the case of *Prosecutor v. Nahimana, Barayagwiz and Ngeze*, distinguishes between statements which discuss ethnic consciousness and those which promote ethnic hatred,[22] giving as an example of the former a personal account of being discriminated against (at the hands of the group against which genocide was later committed).[23] If statements of this sort do motivate listeners to take action, this is, according to the Tribunal, a result, "of the reality conveyed by the words rather than the words themselves."[24] This makes sense conceptually, while also highlighting the problems with the CERD prohibitions on positive statements. In practice, the difference may come down largely to intent rather than content, *per se*. In other words, a distinction may be made between statements whose real intention is positive (to express group pride or consciousness) and those which really aim to denigrate other groups.[25]

Although far less explicit, some of the European Court of Human Rights decisions may be taken to stand for similar propositions. In *Incal v. Turkey*, for example, the Court recognised that the impugned statements appealed to Kurds, urging them to band together to defend their rights. But it held that there was nothing in the text that incited to "violence, hostility or hatred between citizens."[26] This suggests that if racially motivated acts had followed the statement, the fault would not have been attributed to the speaker, as in the *Nahimana* example, but to the context. It may be noted that ECRI's Recommendation 7 calls for the criminalisation of a superiority ideology only where it is expressed with a racist aim.[27]

---

[21] Mahalic, D. and Mahalic, J. "The Limitation Provisions of the International Convention on the Elimination of All Forms of Racial Discrimination" (1987) 9 *Human Rights Quarterly* 74, p. 95.
[22] 3 December 2003, ICTR-99-52-T (Trial Chamber), para. 1020.
[23] Para. 1019.
[24] Para 1020.
[25] There remains, of course, the further problem that CERD does not require positive statements to constitute incitement.
[26] 9 June 1998, Application No. 22678/93, para. 50.
[27] Provision 18(d).

Mahalic and Mahalic also note that differences exist among Committee members regarding academic, scientific or serious debate on matters concerning race, with some suggesting that conclusions which support racial superiority should not be disseminated and others suggesting that they may be disseminated but only if accompanied by arguments discrediting racism and a health warning as to the fallibilities of the study or argument in question.[28] It is hard to see how the latter could be read into the language of Article 4(a)(1).

There would thus appear to be three fundamental incompatibilities between the hate speech regimes found in CERD and the ICCPR, based on the direct language used in their respective provisions. The first is that CERD bans certain statements even in the absence of any intention to incite to or promote hatred. The second is that CERD bans certain statements which do not incite to any particular result, such as hatred, discrimination or violence. Finally, CERD bans certain positive statements of superiority, without requiring that these statements include a negative element relating to the inferiority of other groups. None of these standards are consistent with the rules found in Articles 19 and 20 of the ICCPR.

[28] Pp. 95-96.