Defining and diminishing hate speech

Susan Benesch
Hate speech is the special scourge of minorities and indigenous peoples: like a disease that afflicts only certain populations, it can cause some people to suffer greatly, while others remain unaware and unsympathetic. It gives rise to both psychological and physical harm, and affects a variety of minority and indigenous communities. For example, hate speech has recently been followed by violent attacks against Coptic Christians in Egypt, Muslims in Burma and immigrants in Greece. In those and other parts of the world, hate speech is thriving, nurtured by coinciding factors: economic hardship, large-scale migration, competition between groups for political power after the fall of repressive central regimes, and the ease of expressing hatred online.

The growth of hate speech has inspired alarm in diverse quarters, because it can cause or inspire serious harm in several ways. It directly affects its targets – the people it purports to describe – by frightening, offending, humiliating or denigrating them. This often has the secondary effect of silencing them, by means of fear. Speech can also harm indirectly (but no less severely) by inciting, or pitting members of one group of people against another. Hatred, discrimination and dehumanization are steps in a process that can lead to violence. In Kenya, for example, there is consensus that months of hate speech before the 2007 presidential election contributed to severe violence that broke out when the results were disputed. Since then Kenya has formed a new national agency, the National Cohesion and Integration Commission, charged with reducing hate speech, among its other duties.¹

At the regional and international levels, too, there are new efforts to respond to hate speech. The Council of Europe will soon complete a two-year project called the ‘No Hate Speech Movement’, focused on youth and on what they read, write and hear online. Frank LaRue, the UN Special Rapporteur on freedom of opinion and expression, devoted his 2012 annual report to the ‘increasingly visible’ manifestations of hate speech. He noted several cases in which killings were linked to incitement, such as the murders of Ahmadiyya community leaders in Pakistan after a television broadcast in which two maulanas said the Ahmadiyya deserved to die; violence against members of the Catholic and Jewish communities in Venezuela after incitement against them; and incitement to racism in Israel against the Arab population in Israel, and against Jews in the occupied Palestinian territory.

Several converging reasons help to explain why hate speech is on the rise in so many countries – and also in the transnational virtual space of the internet. First, migration and refugee flows have established new minority communities at the same time as economic and political changes have increased the tendency to stigmatize them. In countries as varied as Greece, South Africa, Côte d’Ivoire and Japan, economic hardship is blamed on minorities, especially those who are viewed by some of the majority population as foreigners because their ancestors immigrated, even though the present-day members of minority groups are native-born. In other cases, political leaders scapegoat minorities to galvanize their supporters, or leaders of ethnic or religious groups jockey for political power by pitting their followers against one another.

Hate speech is disseminated by many means, including the traditional soapbox and bullhorn, graffiti, speeches recorded on CDs and digital communications. It is rife on the internet and social media because some feel free to express their hatred and anger there, even when they would not do so in similarly public settings offline. Online, hateful speakers encourage and incite one another, and cause extra pain to their targets, who are often now privy to hate speech that they would not have seen or heard if it were shared among haters offline, as was more common in the past.² Online platforms such as YouTube, Facebook and Twitter have contributed to ‘a sudden and rapidly increasing wave of bigotry-spewing videos, hate-oriented affinity groups, racist online commentary, and images encouraging violence against the helpless and minorities – blacks, Asians, Latinos, gays, women, Muslims, Jews – across the Internet and around the world’.³ On social media, in particular, civility can quickly disappear, and the most vicious speech becomes commonplace. An activist in Burma compared online spaces to toilet stalls where: ‘people write whatever they want on the walls’.⁴ Inflammatory falsehoods targeting minorities have also become a familiar feature online, where they are powerful
and difficult to refute. For example, in the case of Burma, graphic images of violence in the Central African Republic have been used to argue that Muslims are, as a group, given to savagery.

The almost worldwide explosion in the use of social media is part of a larger change that is not exclusively worrisome: human communities – including members of minority and indigenous, as well as majority, communities – are speaking and listening to one another more than ever before. As United Nations High Commissioner for Human Rights Navi Pillay put it, when launching a two-year study on the prohibition of incitement to hatred in 2011:

‘[T]he enriching variety of backgrounds merging in communities around the world physically or virtually through modern technology also presents a mounting challenge to States as they seek to promote and ensure mutual respect, social harmony, equal opportunity and fairness of treatment to all.’

This technological merging of diverse communities can represent a new opportunity, as people are newly connected across traditional social boundaries – such as those between majority and minority groups – and learn to speak with one another in constructive ways that diminish mistrust and hatred. A small but growing set of efforts to counter hatred online has produced some new ideas and some cause for cautious optimism. They are described briefly below.

Existing law on hate speech
Paradoxically, in spite of the apparent rise in hate speech in many countries, and anxiety about it, there is no consensus on what the term ‘hate speech’ means, either in law or in common parlance; in the words of Kenan Malik, ‘if you look at hate speech laws across the world, there is no consistency about what constitutes hate speech’. This raises serious concerns for freedom of expression, since efforts to restrict hate speech can easily misfire because the term and related law are unclear. Hate speech laws have also been used to attack minorities instead of protecting them – for example, against Roma in Hungary, where anti-Roma hate speech is rife.

In general, hate speech is an expression that denigrates or stigmatizes a person or people based on their membership of a group that is usually but not always immutable, such as an ethnic or religious group. Sometimes other groups, defined by disability or sexual orientation, for example, are included. Speech may express or foment hatred on the basis of any defining feature of a minority or indigenous people, such as ethnicity or religion – and can also denigrate people for another ‘failing’, such as their gender or even their location, as in the case of migrants. When a group is doubly stigmatized in this way, it is known as intersectionality.

This leaves room for many variations, and for unanswered questions, such as: what is hate exactly, and to whose hate does the term ‘hate speech’ refer? The ‘hate’ in hate speech might identify the state of mind of the speaker, or the likely increase of hateful thoughts among a receptive audience, or the terrible and frightening feeling of being hated, on the part of the people whom the speech purports to describe. Indeed the English-language term ‘hate speech’ is relatively new in the literature, as well as scarce in law. It is nearly absent from the texts of Google’s large archive of books published before 1980, but since then its use within the same archive has increased more than tenfold.

Furthermore, the terms ‘hate’ or ‘hatred’, where they are defined at all in law, are usually construed narrowly. For instance, Canada’s criminal code provision against the ‘wilful promotion of hatred’ must be ‘construed as encompassing only the most severe and deeply felt form of opprobrium’, the Canadian Supreme Court found in its landmark case of James Keegstra, a public school teacher who told his students that Jews were an evil people who had ‘created the Holocaust to gain sympathy’.

By itself, hatred cannot and must not be outlawed since it is a state of mind and is not always contemptible, as the legal scholar Robert Post notes with apt reference to Leaves of Grass, in which Walt Whitman exhorted his readers to ‘hate tyrants’, among other good advice. ‘When the law seeks to suppress hate – and hence hate speech – it is not because hate as such ought to be proscribed,’ Post argues. ‘It is instead because the law is intolerant of hatred when it is expressed in particular circumstances. But what
A similar approach is to focus on the harms inflicted by hate speech, and point out which of those harms the law is intended to prevent. In a March 2014 decision in the case of Sangathan v. Union of India, the Supreme Court of India warned against defining hate speech in terms of subjective, individual offence, drawing attention instead to a different harmful consequence, that hate speech can prevent members of minority groups from participating fully in democracy:

‘Hate speech is an effort to marginalise individuals based on their membership in a group. Using expression that exposes the group to hatred, hate speech seeks to delegitimise group members in the eyes of the majority, reducing their social standing and acceptance within society. Hate speech, therefore, rises beyond causing distress to individual group members. It can have a societal impact. Hate speech lays the groundwork for later, broad attacks on vulnerable that can range from discrimination, to ostracism, segregation, deportation, violence and, in the most extreme cases, to genocide. Hate speech also impacts a protected group’s ability to respond to the substantive ideas under debate, thereby placing a serious barrier to their full participation in our democracy.’ (sic)

The idea that hate speech damages democracies by silencing and politically disabling minorities was advanced by Jeremy Waldron in his 2012 book *The Harm in Hate Speech*. Although the United States is notable for its lack of law against hate speech as such – hateful speech is a US crime only if it is likely to lead directly to a serious harm such as ‘imminent lawless action’ – the book was widely discussed. Some descriptions of hate speech define it by naming types of groups that may be targeted. The Council of Europe’s definition of hate speech, for example, adopted in 1997, makes specific reference to Jews, migrants and minorities:

‘[T]he term “hate speech” shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.’

Under nearly all definitions – confusingly to some – it is not hate speech to say or communicate ‘I hate you’ or that one hates a particular political candidate, a political party or its members. The term is widely misunderstood and differently interpreted, raising serious concerns for freedom of expression. During Kenya’s 2013 electoral process, for example, it became common for politicians to accuse one another of hate speech: the accusation became a political weapon and, some argued, it suppressed debate during an election when it was more needed than ever. And in April 2014, a Kenyan government official said she had accused four human rights lawyers of hate speech and fomenting hatred – because they criticized the government’s new policy of rounding up and deporting large numbers of Somalis. In other countries, such as Hungary and Rwanda, laws related to hate speech are used to suppress the grievances and political concerns of minorities and indigenous peoples.

Hate speech laws differ in part because most of them do not use the term ‘hate speech’ at all, referring instead to a variety of acts such as incitement and discrimination or, in Rwanda, the unique offence of ‘ethnic divisionism’ – and a variety of consequences of hate speech, including insult, offence, humiliation and degradation. Laws vary also along other vectors, including the intent of the speaker, the likely effect of the speech, and whether the speech calls for action of some kind. Bhikhu Parekh has illustrated the diversity of national laws with a set of examples:

‘Britain bans abusive, insulting, and threatening speech. Denmark and Canada prohibit speech that is insulting and degrading; and India and Israel ban speech that incites racial and religious hatred and is likely to stir up hostility between groups. In the Netherlands, it is a criminal offence to express publicly views insulting to groups of persons. Australia prohibits speech that offends, insults, humiliates, or intimidates individuals or groups, and some of its states have laws banning racial vilification. Germany goes further, banning speech
that violates the dignity of an individual, implies that he or she is an inferior being, or maliciously degrades or defames a group.'

Some countries have gradually amended their laws prohibiting hate speech, to encompass particular forms of it as they became more prevalent or to prevent misuse of the law. For example, the United Kingdom’s Public Order Act of 1986 identified speech that is ‘threatening, abusive, or insulting’ as an offence if the speaker intended to stir up racial hatred with it, or if, ‘having regard to the circumstances racial hatred is likely to be stirred up thereby’. The Act did not include hatred focused on national, religious or ethnic groups. To partially remedy this omission, the Racial and Religious Hatred Act of 2006 added hate speech against religious minorities. The Act was also amended in 2013, to remove the word ‘insulting’, after a civil society campaign against it. One of the campaigners called it ‘a much-abused catch-all provision where the police could charge anyone for using trivial words that irritated them’.

‘Hate speech’ as such is absent from international human rights treaties, which refer instead to offensive, inciting or discriminatory speech. The Universal Declaration of Human Rights holds in Article 7 that all are entitled to protection against discrimination in violation of the Declaration – and against ‘any incitement to such discrimination’. International law codifies several forms of incitement as offences: incitement to genocide, to violence, to discrimination and to hostility. Article 20(2) of the International Covenant on Civil and Political Rights (ICCPR) states that ‘any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law’. This provision is somewhat unclear, since the distinctions between advocacy and incitement, and hatred and hostility are contested. Article 20 has been incorporated only partially, or not at all, into bodies of national law. The International Convention on the Elimination of All Forms of Racial Discrimination (CERD) addresses hate speech most directly in its Article 4 condemning:

‘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.’

In this regard, hate speech laws can be helpful in tackling impunity and in helping to establish social norms against certain types of speech and, more importantly, against certain types of harm. Such laws cannot be successful without the support of influential public figures, however, nor can they succeed unless it is clear which speech they prohibit and why.

To focus efforts to counter hate speech and to prevent one of its worst apparent consequences – violence – I have described a specific subcategory of hate speech, defined by its capacity to inspire or catalyse group violence, and have named this category ‘dangerous speech’. Across a great variety of languages, countries and historical periods, such language exhibits characteristic patterns or ‘hallmarks’. It frequently compares other human beings to pests or vermin, for example, suggesting that members of minority groups within a country are foreign or that they besmirch, taint or damage the majority group.

Most existing hate speech laws – including international, regional and national ones – are dangerously vague, in ways that are often used to restrict the freedom of speech of minorities, including preventing them from expressing legitimate grievances. Even David Matas, a vigorous proponent of criminalizing hate speech, writes that ‘if anti-hate laws are vague, we should not have them. Vagueness would vitiate the laws, render them useless, and indeed threaten free speech unduly.’ Indeed, the hate speech provision in Canada’s Human Rights Act was struck down in June 2013, though its criminal law provision against hate speech remains law. ‘We don’t want there to be a chill on speech that is controversial but not necessarily hateful,’ Cara Zwibel of the Canadian Civil Liberties Association said when the Human Rights Act’s hate speech clause was struck down.

‘We felt that given the impact that it has on freedom of expression, and given that it hasn’t really proven to be a very effective method for dealing with discrimination, that it shouldn’t be on the books
anymore…. We really encourage countering hateful speech, rather than trying to censor it.”

Counter speech and other alternatives to criminal law

New methods to diminish hate speech – or reduce its impact – without infringing on freedom of expression are emerging. Activists, journalists, clergy, lawyers and others have begun experimenting with such methods in a variety of countries. Technology plays a role in many of these efforts: just as new communications technologies are being used to amplify inflammatory hate speech, they can also be marshalled to prevent and counter it. New technologies are also being employed to detect where hate speech may signal an increased risk of mass violence.

To prevent mass violence, especially in societies at high risk, advocates have begun to experiment with media programming to render audiences less likely to become convinced by inflammatory speech, or to act on it. This is called ‘inoculating’ an audience, following the example of the non-governmental organization Radio la Benevolencija (RLB), which uses the term to refer to its own path-breaking work. RLB, based in Amsterdam and working in several central African countries, has collaborated with the social psychologist and genocide scholar Ervin Staub to develop ‘knowledge tools’ – guides on how to deal with an array of manipulative pressures that move individuals and whole societies to physical and mental harm, and how to resist such pressures. RLB delivers these tools in entertaining programmes such as a radio soap opera called Musekeweya (New Dawn), which has become popular in Rwanda since its launch there in 2004. What particularly distinguishes Musekeweya from other soap operas in Rwanda (or indeed, elsewhere), according to RLB, is that it ‘explicitly deals with the psychology of incitement to hate and violence that leads to mass conflict’. Musekeweya’s impact on its listeners was independently studied by a scholar, Elizabeth Levy Paluck, who described her effort as ‘the first experimental evaluation of a radio program’s impact on intergroup prejudice and conflict in a real world setting’. In her year-long study, Paluck found ‘a pattern of norm and behavior change’ and an increase in empathy, on the part of Musekeweya listeners, for other Rwandans.

In another effort in Kenya in 2012, four episodes of a long-standing Kenyan television comedy/drama called Vioja Mahakamani (Events in the Courtroom) took up the topic of hate speech. In each of four episodes, a Kenyan (or group of Kenyans) stands accused either of making inflammatory hate speech (at a rally or on a printed flyer, for example) or of acting upon it, and the harms of hate speech are discussed as the case goes forward. The episodes’ impact on audiences was independently evaluated by scholars who found that Kenyans who watched the episodes felt better able to identify and to resist incitement.

Inoculation against hate speech takes some time, and therefore should be conducted in advance (just like the more familiar kind of inoculation, against disease) before the risk of violence becomes acute. Especially (but not only, of course) during that acute stage, there is a second type of alternative method for diminishing the force or effectiveness of dangerous speech that I describe with the general term counter speech – or speech to refute hate speech or dangerous speech.

Counter speech may be effective at forestalling the effects of hate speech, including violence. There is some evidence of success when influential or prominent leaders publicly and unequivocally indicate that they disapprove of hate speech – or of violence itself – even though strong counter speech is relatively uncommon. In one example, Norwegian Prime Minister Jens Stoltenberg declared ‘we will answer hatred with love’ after Anders Breivik massacred 77 people, mostly teenagers, in 2011. After the killing of filmmaker Theo van Gogh in 2004, Amsterdam’s mayor Job Cohen spoke out firmly against the angry anti-Muslim rhetoric and sentiment that followed. He ‘initiated the peace script’, as the New York Times later put it, for example by telling the people of his city, ‘An Amsterdamer is murdered. You fight with the pen and, if necessary, in the court. But never take the law in your own hands.’ In the days after van Gogh was killed, revenge attacks against Muslims happened elsewhere in the country, but not in Amsterdam.

According to scholarly research on why Hindu–
Muslim riots and massacres have happened in India at some volatile times and places, but not in others, influential figures have been able to thwart violence by publicly withdrawing their support for it, even where extremist parties controlled relevant state governments. Likewise the King of Denmark and Danish political leaders have been credited with helping to save the lives of nearly all of Denmark’s Jews during the Holocaust, in part by consistently speaking of them as part of the same national community as other, non-Jewish Danes. ‘I considered our own Jews to be Danish citizens,’ King Christian was quoted as writing, in the Danish historian Bo Lidegaard’s 2013 book Countrymen, ‘and the Germans could not touch them. The prime minister shared my view and added that there could be no question about that.’

Counter speech can also be effective when it comes from a wide variety of sources, speaking in unison. Kenya produced an example of this in the weeks and months before its presidential election in March 2013, the country’s first since inflammatory speech and severe violence accompanied the attempted election of 2007.

Thought leaders of all kinds called on Kenyans to forsake hatred and violence. Ecumenical groups of clerics appeared on billboards and on the radio, calling for peace. Graffiti artists covered walls and fences with anti-hate murals. Football stars recorded brief public service announcements, appealing directly to young men like themselves to remain calm. Amid this unprecedented volume of anti-hatred ‘peace propaganda’, the election went forward with only one episode of serious violence, by a local extremist group which tried to disrupt the election by attacking polling places in the city of Mombasa.

In Burma, an innovative counter speech campaign began in April 2014, after months in
which politicians, Buddhist monks and other influential leaders produced increasingly violent, vicious hate speech against Muslims. The ‘panzagar’ or ‘flower language’ campaign is led by Nay Phone Latt, a blogger and free speech advocate in Burma who was arrested for reporting on anti-government protests in 2007 and sentenced to 20 years in prison. He was released after four years, and is now working with other human rights activists on Panzagar, which uses the symbol or meme of a person holding a flower in his or her mouth, to counter hate speech. The Panzagar message is being distributed in many ways: with a song, on flyers and on a page on Facebook, the social media platform that is overwhelmingly the most used in Burma. In an April 2014 interview with The Irrawaddy newspaper, Nay Phone Latt explained why he is now dedicating so much of his formidable energy to countering hate speech:

‘If people hate each other, a place will not be safe to live. I worry about that most for our society. In some places, although they are not fighting, hate exists within their heart because they have poured poison into their heart for a long time [through hate speech]. It can explode in anytime.’

Endnotes

1 Kenya’s National Cohesion and Integration Commission was first established by the National Cohesion and Integration Act of 2008, passed in the wake of the post-election violence of 2007–8, in which more than 1,000 people were killed and about 500,000 displaced. The Commission was made statutory by Kenya’s new National Constitution, passed in 2010. For a description of other efforts to counter hate speech during Kenya’s subsequent presidential election in 2013, see Susan Benesch, ‘Countering dangerous speech to prevent mass violence during Kenya’s 2013 election’, retrieved April 2014, http://voicesthatpoison.org/kenya-2013/.


3 Ibid., p. 11.


7 This study was conducted by the author, using the NGram utility for Google Books.


14 Another forthcoming RLB program has this memorable, brilliant title: *Hate: A Course in Ten Easy Lessons*.


16 Basu, A., ‘When local riots are not merely local’,
Minority Rights Group International

Minority Rights Group International (MRG) is a non-governmental organization (NGO) working to secure the rights of ethnic, religious and linguistic minorities and indigenous peoples worldwide, and to promote cooperation and understanding between communities.

Our activities are focused on international advocacy, training, publishing and outreach. We are guided by the needs expressed by our worldwide partner network of organizations which represent minority and indigenous peoples.

MRG works with over 150 organizations in nearly 50 countries. Our governing Council, which meets twice a year, has members from nine different countries. MRG has consultative status with the United Nations Economic and Social Council (ECOSOC), observer status with the African Commission on Human and People’s Rights, and is registered with the Organization of American States.

MRG is registered as a charity and a company limited by guarantee under English law. Registered charity no. 282305, limited company no. 1544957.

Discover us online:

**MRG website**
Visit our website for news, publications and more information about MRG’s work:
www.minorityrights.org

**Minority Voices Newsroom**
An online news portal that allows minority and indigenous communities to upload multimedia content and share their stories:
www.minorityvoices.org

**Peoples under Threat**
MRG’s annual ranking showing countries most at risk of mass killing is now available as an online map:
www.peoplesunderthreat.org

**World Directory of Minorities and Indigenous Peoples**
The internet’s leading information resource on minorities around the globe:
www.minorityrights.org/Directory
Across the world, minorities and indigenous peoples are disproportionately exposed to hatred. From intimidation and verbal abuse to targeted violence and mass killing, this hatred often reflects and reinforces existing patterns of exclusion. The impacts also extend beyond the immediate effects on individual victims to affect entire communities – in the process further marginalizing them from basic services, participation and other rights.

This year’s edition of *State of the World’s Minorities and Indigenous Peoples* highlights how hate speech and hate crime, though frequently unreported or unacknowledged, continue to impact on every aspect of their lives. The volume also documents many of the initiatives being taken to promote positive change and the different ways that governments, civil society and communities can strengthen protections for minorities and indigenous peoples.