THE PROBLEM WITH HATE SPEECH LAWS

By Jacob Mchangama

On August 19, 2014, the city court of Malmö, Sweden, convicted Swedish "shock-artist" Dan Park of "agitation against a national or ethnic group" under Sweden's anti-hate speech laws and sentenced him to six months imprisonment. The court also ordered that his art works, which had been seized by police officers, be destroyed. This was not the first time Park fell afoul of the law, and having already served three months in prison, Park has spent more time locked up than even Chinese artist Ai Wei Wei.

Park's alleged crime was exhibiting a number of provocative and "politically incorrect" art works, including one with the words "hang on Afrofobians" and a picture of a number of prominent African-Swedish anti-racist activists with ropes around their necks as well as holding a placard with the words "Gypsy crime is a good thing." In its decision, the city court of Malmö explicitly referred to the case law of the European Court of Human Rights (hereafter, the "Court"), insisting that Park "had an obligation to avoid being gratuitously offensive to others." Park's conviction is a good example of how prohibitions against "hate speech" are firmly entrenched in international human rights law and—often as a result thereof—national laws throughout the world.

This may explain why mainstream human rights organizations have not condemned the artist's imprisonment and the destruction of his work. In fact, Amnesty International chapters in both Sweden and Denmark supported Park's conviction (Jyllands-Posten 2014). Yet, this essay will argue that there are strong reasons to question the inclusion of hate speech bans in international human rights law. First, the drafting history of international anti-hate speech laws shows that such laws are a legacy of totalitarian states aimed at abusing human rights rather than strengthening tolerance. Second, the applicable standards are conflicting, impossible to reconcile with the principle of legal certainty inherent in the rule of law, and prone to abuses that undermine critically important freedoms of speech—especially political speech. Third, laws against hate speech and against "offense" are tools in the hands of those who would severely restrict religious freedom. Finally, proponents of hate speech bans have yet to demonstrate convincingly any link between such bans and social peace and tolerance.

The History of Hate Speech Bans in Human Rights Law

It is a little known fact that hate speech bans in international human rights law were first introduced by authoritarian states. In fact, during the cold war, the battle over the limits of free

Jacob Mchangama is founder and Director of The Freedom Rights Project, a human rights organization dedicated to restoring liberty as the heart of human rights. He is also the founder and executive director of Justitia, a new Copenhagen based think tank focusing on human rights and the rule of law. He has written and commented on human rights in international media including *Foreign Policy, Foreign Affairs, The Economist, BBC World, Wall Street Journal Europe,* and *The Times.*

speech in human rights law became a proxy for the conflict in the United Nations between liberal democracies and communist states.

The (nonbinding) Universal Declaration of Human Rights (UDHR) adopted in 1948 does not include an explicit duty to prohibit hate speech. Article 19 simply secures "freedom of opinion and expression." However, the drafting history reveals frequent discussion of hate speech restrictions. The drafters faced the challenge of whether and to what extent freedom of expression should be a principle under which intolerance is tolerated. The Soviet Union was the primary advocate for hate speech restrictions, whereas the vast majority of Western democracies, led by the United States and the United Kingdom, sought to guarantee broad protection for free speech and avoid any explicit obligation upon states to restrict this right.

Unlike the UDHR, the International Covenant on Civil and Political Rights (ICCPR) is a legally binding human rights convention, currently ratified by 167 states. The ICCPR was adopted in 1966 and includes a right to freedom of expression in Article 19, but also an obligation to prohibit hate speech in Article 20 (2): "Any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence shall be prohibited by law."

The adoption of Article 20 was highly controversial and was preceded by heated negotiations. The first draft was limited to the prohibition of "any advocacy of national, racial, or religious hostility that constitutes an incitement to violence." However, a number of countries led by the Soviet Union sought a broader prohibition against "incitement to hatred." Proponents of hate speech prohibitions justified Article 20 by referring to World War II and the Holocaust. Over time, colonialism and apartheid were also used as justifications for prohibiting racial and religious hatred (Farrior 1996). Article 20's opponents argued that such restrictions do not belong in a human rights convention; more specifically, the terms "hatred" and "hostility" are vague and risk arbitrarily undermining freedom of expression. Eleanor Roosevelt, then chairman of the UN's Commission on Human Rights, found the language "extremely dangerous" and warned against provisions "likely to be exploited by totalitarian states for the purpose of rendering the other articles null and void." She also feared that the provision "would encourage governments to punish all criticism under the guise of protecting against religious or national hostility" (cited in Mchangama 2011).

The General Assembly adopted Article 20 with 52 votes in favor, 19 against, and 12 abstentions. Those in favor consisted primarily of the communist states of Eastern Europe, as well as non-Western countries with very questionable human rights records such as Saudi Arabia, Haiti, Sudan, and Thailand. The 19 countries that voted against the provision included most Western liberal democracies of the time as well as Ecuador, Uruguay, Japan, Malaysia, and Turkey. European opposition toward a hate speech ban in the ICCPR mirrored these countries' rejection of Turkish and Greek proposals to insert such languages in the European Convention on Human Rights adopted in 1950 (Mchangama, forthcoming).[1]

While the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) was adopted in 1965, prior to the ICCPR, most of its provisions were drafted after those in the ICCPR. Article 4(a) of ICERD includes an obligation to "declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, [and] incitement to racial discrimination." This provision proved a major point of contention between liberal democracies and communist states. The latter were now also supported by newly decolonized states that saw the ICERD as an important tool against apartheid and the yoke of colonialism, which they had just escaped (Schwelb 1966). The Colombian representative offered the strongest criticism against this provision, insisting that this clause would be

> a throwback to the past, since punishing ideas, whatever they may be, is to aid and abet tyranny and leads to the abuse of power ... [A]s far as democracy is

> concerned, ideas should be fought with ideas and reasons ... not by ... prison, exile, confiscation or fines.

In the end, Article 4 (a) was softened by the insertion of a clause requiring "due regard" to the rights in the UDHR, including freedom of expression.

The Proliferation of Conflicting Standards

Since the adoption of ICCPR, Article 20, and ICERD, Article 4, democratic resistance to hate speech bans has generally been replaced by enthusiasm, and hate speech bans have proliferated at the national and regional levels. At the heart of this shift is the belief that social peace in an increasingly multiculturalist Europe requires certain restrictions on expressions aimed at racial, ethnic, and religious (and recently also sexual) minorities. For instance, the Council of Europe's 2003 Additional Protocol to the Convention on Cybercrime is aimed at criminalizing intentional "acts of a racist and xenophobic nature committed through computer systems." It defines such acts as

> any written material, any image, or any other representation of ideas or theories which advocates, promotes, or incites hatred, discrimination, or violence against any individual or group of individuals based on race, color, descent, or national or ethnic origin, as well as religion if used as a pretext for any of these factors.

The Protocol even includes an obligation to criminalize "insulting publicly" a person for belonging to a particular racial, ethnic, etc. group.

The EU Framework Decision on Combating Racism and Xenophobia from 2008 obliges all EU member states to criminalize "intentional conduct" aimed at "publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, color, religion, descent, or national or ethnic origin."

The absence of any clear and generally accepted definition of hate speech contributes to the confusion on the permissible limits of speech. The question of permissible limits has generated substantial discussion and controversy at the international level due to fundamental differences among states over the relationship between freedom of expression and impermissible hate speech—and religious hate speech in particular.

As explained earlier, the ECHR does not require a ban on hate speech. However, the Court has long excluded hate speech from the freedom of expression protections of Article 10. In *Gündüz v. Turkey* (2003) the Court held:

> [T]here can be no doubt that concrete expressions constituting hate speech, which may be insulting to particular individuals or groups, are not protected by Article 10 of the Convention.

and that:

> [T]olerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society. That being so, as a matter of principle it may be considered necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance.

While all of the aforementioned definitions overlap, they are not consistent and some are more encompassing than others. For instance, the EU Framework Decision's requirement of intent and incitement to hatred would seem to constitute a significantly higher threshold than both the Court's acceptance of the criminalization of expressions "which may be insulting to particular individuals or groups" and "all forms of expression which spread, incite, promote, or justify hatred." The intent and incitement standard is also narrower than the Optional Protocol to the Convention on Cybercrime's broad prohibition on "insulting publicly" and ICERD's prohibition of "all

dissemination of ideas based on racial superiority or hatred." The Framework Decision's standard is in fact more closely aligned with ICCPR Article 20(2), which requires "incitement," though it does not explicitly require intent.

Moreover, the Court has never defined "hate speech" and does not always distinguish between "incitement" and "offense." In fact, in a number of cases, the Court has accepted even blasphemy laws as permissible limits on freedom of expression (e.g. *Otto Preminger v. Austria* (1994) and *I.A. v. Turkey* (2005)). (For more on relevant European case law and legal standards see report by the Venice Commission 2010.) In 2012, the Court went so far as to state that the right to privacy includes a positive obligation on states to prohibit "negative stereotypes," which must be balanced against the right to freedom of expression (*Aksu v. Turkey* (2012)). This decision clashes with the interpretation of the UN Human Rights Committee, which in General Comment 34 insists that both blasphemy and memorial (such as Holocaust denial) laws would violate ICCPR, Article 19, and that all exceptions to freedom of expression (including hate speech bans) must be convincingly established and narrowly defined.

The lack of a coherent approach to cases of "extreme speech" has resulted in numerous problematic convictions, including that of a French mayor who advocated boycotting of Israel (*Willem v. France* (2009)), a Belgian politician critical of immigration from North Africa (*Ferret v. Belgium* (2009)), and a private individual who, in a personal letter, denied the culpability of Hitler in the Holocaust (*Witzsch v. Germany* (2005)). Moreover, there is evidence of hate speech bans being abused to curb political speech. During Zimbabwe's 2013 elections, for instance, police confiscated radios, alleging that they were used to spread "hate speech" (IFEX 2013). The radios were an important means of receiving news from non-state controlled media. The Council of Europe's Human Rights Commissioner has advocated expanding existing hate speech bans to cover gender-based hate speech as well as a Europe-wide ban against Holocaust denial (www.coe.int, 06/03–2014). These examples demonstrate the danger of internationalizing hate speech bans in human rights law so presciently warned against by Eleanor Roosevelt during the debate on ICCPR Article 20 some six decades ago.

Hate Speech and Religion

The intersection between religion and hate speech is particularly interesting and important. In *Norwood v. United Kingdom* (2004), a British man had been convicted for placing a controversial poster in his apartment window shortly after the 9/11 attacks. The poster depicted the World Trade Center towers in flames, a star and crescent in a prohibition sign, and the words "Protect the British People: Islam out of Britain." The Court held that the poster constituted a

> public expression of attack on all Muslims … Such a general, vehement attack against a religious group, linking the group as a whole with a grave act of terrorism, is incompatible with the values proclaimed and guaranteed by the Convention, notably tolerance, social peace, and non-discrimination.

While there is no doubt that the poster was provocative and would offend many Muslims, it is less clear whether the poster constituted an attack on Muslims rather than an attack on Islam. Categorizing the poster as incitement to hatred against Muslims requires ascribing the worst possible motive to the speaker.

If we accept that scathing criticism of religions constitutes "hate speech," one could presumably also prohibit the following statement:

> Of all the systems of religion that ever were invented, there is none more derogatory to the Almighty, more unedifying to man, more repugnant to reason, and more contradictory in itself, than this thing called Christianity. Too absurd for belief, too impossible to convince, and too inconsistent for practice, it renders the heart torpid, or produces only atheists and fanatics.

But if this statement was prohibited, people who quoted Tom Paine's (1795) *The Age of Reason* could be convicted. Yet *The Age of Reason* is considered an enlightenment classic rather than hate speech, despite its many attacks on Christianity, Judaism, and Islam, including the charge that these religions are "human inventions, set up to terrify and enslave mankind, and monopolize power and profit."

Defining religious hate speech is of particular importance for freedom of expression since a number of states have made a concerted effort to significantly narrow the limits of permissible speech in the sphere of religion. For more than a decade, the Organization of Islamic Cooperation, through successive resolutions at the UN Human Rights Council (HRC) and General Assembly, has pushed for a ban on "defamation of religion." In 2011, a compromise was reached with HRC Resolution 16/18, which omitted any reference to "defamation of religion." The so-called Istanbul Process, a series of intergovernmental meetings, has been preoccupied with how to interpret and implement Resolution 16/18. But it has long been clear that the consensus on 16/18 exists mostly on paper.

On February 26, 2013, the US representative in the HRC, Dr Esther Brimmer, hailed Resolution 16/18 as "a remarkable achievement" and applauded the leadership of Turkey and Pakistan as well as "the support of the OIC Secretary-General." According to Dr Brimmer, "The international consensus on this issue offers a practical and effective means to fight intolerance, while avoiding the false choice of restricting the complementary and mutually-dependent freedoms of religion and expression." But on the previous day, the Pakistani representative, speaking on behalf of the OIC, stated that:

> [T]here are emerging challenges and issues which need to be addressed by international human rights law. … Negative stereotyping or defamation of religions is a contemporary manifestation of religious hatred, discrimination and xenophobia. While the freedom of expression is sacrosanct, it must not be exploited to incite hatred against any religion and violence against its followers.

In other words, the OIC explicitly equated "defamation of religion," a broad and nebulous category including religious satire and criticism, with advocacy of religious hatred, prohibited in Article 20 of the ICCPR. That should not come as a surprise. On numerous occasions subsequently, the OIC has referred to 16/18 as including an expansive interpretation of the prohibition against advocacy of religious hatred, which would cover Danish and French newspapers' publication of cartoons depicting the prophet Mohammed. Thus at the 3d meeting of the So-Called Istanbul Process on 20 June 2013, the Secretary General of the OIC praised a joint statement issued by the OIC, Arab League, the African Union, and the European Union following the release of the film "Innocence of Muslims" on Youtube. The Secretary General highlighted that "[R] esolution 16/18 has indeed been helpful" to form a consensus and that:

> The joint statement condemned a clear act of advocacy to religious hatred that constituted incitement to hostility and violence. It also emphasized the need to respect believers' legitimate and objective sensitivities with regard to the sanctity of religious figures and symbols.

Defining "advocacy of hatred" and "incitement to hostility and violence", this broadly amounts to smuggling a blasphemy ban in through the back door. And OIC states would not be the only ones to benefit from such an expansive interpretation of hate speech laws: In 2005, the director of the Sakharov museum in Moscow was convicted for "religious hatred" after having staged the "Caution! Religion" exhibition, which included artworks aimed at discussing the role of religion in politics (Human Rights Watch 2005). The exhibition was attacked and vandalized by Orthodox extremists, yet after pressure from politicians and the Orthodox Church, the vandals were released and the museum director charged and convicted.

Even in liberal democracies, "scope creep" from blasphemy standards to hate speech bans is an inherent danger. When the city court of Malmö convicted Dan Park and insisted that he "had an obligation to avoid being gratuitously offensive to others," it cited the Court's decision in *Otto Preminger v. Austria* in which it found Austrian authorities' seizure of "blasphemous" film consistent with the right to freedom of expression in the ECHR. But the entire relevant passage of *Otto Preminger* reads as follows:

> [W]hoever exercises the rights and freedoms enshrined in the first paragraph of that Article (art.10) undertakes "duties and responsibilities." Amongst them—in the context of religious opinions and beliefs—may legitimately be included an obligation to avoid as far as possible expressions that are gratuitously offensive to others and thus an infringement of their rights, and which therefore do not contribute to any form of public debate capable of furthering progress in human affairs.

As is evident from above, the duty to refrain from "gratuitous offense"—itself a deeply subjective and troublesome standard—applies in cases relating to religious sensitivities, not hate speech. But by using the much lower threshold of unqualified "offense," the City Court of Malmö significantly widened the scope of Sweden's hate speech laws with reference to international human rights standards. This is exactly the strategy envisaged by the OIC when it targeted the "defamation of religion" through hate speech bans.

The Missing Link between Hate Speech and Hate Crimes

It could be argued that legal uncertainty and the risk of abuse are costs worth paying if the result is prevention of social unrest and intolerance that may lead to the type of race- or religion-based violence witnessed in Rwanda, the former Yugoslavia, or even the Holocaust. Yet there exists no evidence that hate speech involves a measurable harm or that speech restrictions are a useful tool in curbing such harms. In fact, recent research shows that countries that restrict freedom of expression through blasphemy laws experience *more* religiously motivated social unrest and violence than countries that do not enforce such laws (Pew 2011).

Consider also the American case, where robust protections of freedom of speech and religious freedom have coincided with increases in *tolerance*, not intolerance. American attitudes toward interracial marriages have shown a dramatic shift towards acceptance from 4 percent in 1958 to 87 percent approval in 2013 (Gallup 2013). Similarly, an Anti-Defamation League (ADL) survey shows that in 2013, 12 percent of Americans harbored anti-Semitic attitudes (as defined by the ADL), as opposed to 29 percent in 1964 and 15 percent in 2012. The steady drop in American anti-Semitism and an increase in the levels of racial tolerance and interracial marriages has taken place concurrently with the US Supreme Court's strengthening of the First Amendment's protection of free speech (see e.g. *Brandenburg v. Ohio* 395 U.S. 444 (1969).

In the United States, mere expressions of hatred toward Jews or other groups are generally protected as free speech unless they constitute *incitement to imminent violence*. This standard has led to, for example, First Amendment protection for Neo-Nazis marching through Skokie, Illinois, a town near Chicago inhabited by one of the largest populations of Holocaust survivors outside Israel.

In Europe, with its ubiquitous hate speech laws, the situation is quite different. In October 2013, the European Union Agency for Fundamental Rights issued a report based on a survey of more than 5000 European Jews. Seventy-six percent of those surveyed responded that anti-Semitism has become worse in their respective countries within the last five years (European Agency for Fundamental Rights 2013). This pattern fits with ADL's surveys of anti-Jewish sentiments; its 2012 survey revealed that all but one of 10 European countries studied had experienced an increase in anti-Semitic attitudes (as defined by the ADL) since their previous survey in 2009 (Anti-Defamation League 2012).

These statistics do not allow us to draw firm conclusions about correlation or causation with free speech, nor can we easily compare the United States with Europe. However, if we accept that the burden of proof for restricting free speech rests on those who want to limit this fundamental right, then some prima facie evidence is needed.

Those who favor hate speech bans may also consider whether such bans are not only illiberal and ineffectual, but also counterproductive—at least in liberal democracies where speech cannot be suppressed through elaborate censorship schemes or lengthy prison sentences (though the Dan Park case might signal a worrying shift). Take for example the Dutch politician Geert Wilders, infamous for his Muslim-baiting and harsh attacks against Islam. In January 2010, he was charged with violating the Netherland's hate speech laws, which in turn created a storm of media attention. In the June 2010 parliamentary elections, Geert Wilders's party won a total of 24 seats, an increase of 15, to become the third largest party in the Netherlands. Shortly after the election, Wilders was acquitted. Similarly, when the French Interior Minister banned French comedian Diedonné from performing live shows in France because of the anti-Semitic content of the shows, Dieudonné made a Youtube video mocking the French minister. The video attracted more than 2 million views. Both Wilders and Dieudonné have thus used hate speech bans as a megaphone for their messages, reaching millions of people while portraying themselves as free speech martyrs suppressed by the establishment. These applications of hate speech bans are hardly examples of "best practices" in the important fight against racism and intolerance.

THOSE WHO FAVOR HATE SPEECH BANS MAY ALSO CONSIDER WHETHER SUCH BANS ARE NOT ONLY ILLIBERAL AND INEFFECTUAL, BUT ALSO COUNTERPRODUCTIVE

All these arguments should prompt proponents of hate speech laws to revisit their support for legal instruments that are the natural tools of dictatorships, but serve as a Damoclean sword in liberal democracies. ❖

1. However, these same states accepted that the convention should limit totalitarian movements from gaining a foothold in Europe.

References

Anti-Defamation League. 2012. "Attitudes Toward Jews In Ten European Countries." Anti-Defamation League Website, March. http://www.adl.org/assets/pdf/israel-international/adl_anti-semitism_presentation_february_2012.pdf.

Council of Europe Commissioner for Human Rights. 2014a. "Europe Still Haunted by Anti-Semitism." Commissioner's Human Rights Comments, January 23. http://www.coe.int/en/web/commissioner/-/europe-still-haunted-by-antisemiti-1.

Council of Europe Commissioner for Human Rights. 2014b. "Hate Speech Against Women Should be Specifically Tackled." Commissioner's Human Rights Comments, March 6. http://www.coe.int/en/web/commissioner/-/hate-speech-against-women-should-be-specifically-tackl-1.

European Agency for Fundamental Rights. 2013. "Discrimination and Hate Crime Against Jews in EU Member States: Experiences and Perceptions of Anti-Semitism." http://fra.europa.eu/sites/default/files/fra-2013-discrimination-hate-crime-against-jews-eu-member-states-0_en.pdf.

Farrior, Stephanie. 1996. "Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech." *Berkeley Journal of International Law.* Int'l Law 1. http://scholarship.law.berkeley.edu/bjil/vol14/iss1/1.

Gallup. 2013. "In U.S., 87% Approve of Black-White Marriage, vs. 4% in 1958." Gallup Website, July 25. http://www.gallup.com/poll/163697/approve-marriage-blacks-whites.aspx.

Human Rights Watch. 2005. "Russia: Art Conviction Undermines Free Expression." Human Rights Watch Website, March 28. http://www.hrw.org/news/2005/03/27/russia-art-conviction-undermines-free-expression.

IFEX. 2013. "Police in Zimbabwe ban 'specifically designed radios,' claiming used for hate speech." IFEX, February 21. https://www.ifex.org/zimbabwe/2013/02/21/zimbabwe_alert_police/.

Jyllands-Posten. 2014. "Fængsling af svensk kunstner udløser hård dansk kritik," August 29. http://www.jyllands-posten.dk/protected/premium/kultur/ECE6979515/F%C3%A6ngsling+af+svensk+kunstner+udl%C3%B8ser+h%C3%A5rd+dansk+kritik/.

Mchangama, Jacob. 2011. "The Sordid Origin of Hate-Speech Laws." *Policy Review* 170. http://www.hoover.org/research/sordid-origin-hate-speech-laws.

Mchangama, Jacob. Forthcoming. "The European Convention on Human Rights and 'Extreme Speech': What's Left of the Right to 'Offend, Shock, and Disturb'?

Paine, Thomas. 1795. *Age of Reason.*

Pew Forum. 2011. "Rising Restrictions on Religion – One-third of the world's population experiences an increase." Pew Forum, August 9. http://www.pewforum.org/2011/08/09/rising-restrictions-on-religion2/.

Schwelb, Egon. 1966. "The International Convention on the Elimination of All Racial Discrimination." *International and Comparative Law Quarterly* 15 (4): 996–1068.

Venice Commission. 2010. "Blasphemy, Insult and Hatred: Finding Answers in a Democratic Society." Science and Technique of Democracy, No. 47. http://www.venice.coe.int/webforms/documents/?pdf=CDL-STD%282010%29047-e.

http://dx.doi.org/10.1080/15570274.2015.1005919