



Hate speech or free speech: an ethical dilemma?

Nina Gorenc

To cite this article: Nina Gorenc (2022) Hate speech or free speech: an ethical dilemma?, International Review of Sociology, 32:3, 413-425, DOI: [10.1080/03906701.2022.2133406](https://doi.org/10.1080/03906701.2022.2133406)

To link to this article: <https://doi.org/10.1080/03906701.2022.2133406>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 34365



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 13 View citing articles [↗](#)

Hate speech or free speech: an ethical dilemma?

Nina Gorenc 

Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

ABSTRACT

Freedom of expression is the cornerstone and the warrant of democracy, but like all other rights and freedoms we enjoy, also the right to free speech has its limits. Hate speech is one of the most resilient manifestations of cyberviolence, and is not to be equalled with free speech. This is partly related to our perception of the freedom of expression that has substantially changed through time; from the past focus on being able to freely express your opinion, to question, dissent and challenge the government, to today's need to limit these rights in order to protect the targets of hate speech. The article first looks at how hate speech is defined at the level of the EU and at the problems connected with its regulation. It continues with an overview of relevant literature and research on the topic, and concludes by analyzing and interpreting the results of an online survey conducted among Slovenian university students in February 2022. Our goal is to understand the reasons behind the rise of hate speech and reactions to it; from its acceptability and underestimation to indifference and ignorance.

KEYWORDS

Freedom of expression; free speech; hate speech; censorship; social media

Introduction

Hate speech has become some sort of a controversial buzz word of today, with many advocates of its regulation on one side, and many opponents on the other. Controversy of this kind is aggravated by the fact that by limiting and regulating hate speech, we erode the freedom of expression and make it a zero-sum game. Furthermore, there is no agreement on what precisely constitutes hate speech, in particular the speech that should be banned and punishable by law. Despite the fact that it can by no means be considered a new phenomenon, there is yet no universal definition of hate speech. However, different international covenants and documents have enshrined the principle of the protection of human dignity already in the 1960s by stating, among other, that 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 are declared an offence punishable by law (United Nations, 1964). Article 20 of International Covenant on Civil and Political Rights (United Nations, 1966) states that 'any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law'.

CONTACT Nina Gorenc  nina.gorenc@fdv.uni-lj.si  Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, Ljubljana 1000, Slovenia

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

The European Union was founded on the values enshrined in Article 2 of the Treaty on the Functioning of the European Union (2012), which Member States committed to respect and promote. These are human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. The EU must aim to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, and ‘take appropriate action to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation’. The EU must

ensure a high level of security through measures to prevent and combat crime, racism and xenophobia, and through measures for coordination and cooperation between police and judicial authorities and other competent authorities, as well as through the mutual recognition of judgments in criminal matters,

and in order to achieve this goal, the EU may adopt measures to approximate criminal laws (Treaty on the Functioning of the European Union, 2012). The European Court of Justice stated that the European Community and the European Court of Justice (ECJ) protect human rights as ‘general principles of law’, and the Community court extended the protection to those rights which arise from the ‘constitutional traditions common to the Member States. When defining these common constitutional traditions, the ECJ drew inspiration from the international conventions signed by all the member states, and regarded the European Convention on Human Rights as the main common source (Bayer & Bard, 2020, p. 49).

The EU law defines illegal hate speech as public incitement to violence or hatred on the basis of certain characteristics, including race, colour, religion, descent and national or ethnic origin. While the Framework Decision on combatting racism and xenophobia (2008) covers only racist and xenophobic speech, the majority of Member States have extended their national laws to other grounds such as sexual orientation, gender identity and disability. The EU provides some non-binding definitions of hate speech, e.g. the Council of Europe (1997) defines hate speech as

all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism 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,

while the European Commission against Racism and Intolerance (2015) defines it as

the use of one or more particular forms of expression, namely the advocacy, promotion or incitement of the denigration, hatred or vilification of a person or a group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat of such person or persons and any justification of all these forms of expression – that is based on a non-exhaustive list of personal characteristics or status that includes ‘race’, colour, language, religion or belief, nationality or national or ethnic origin, as well as descent, age, disability, sex, gender, gender identity and sexual orientation.

The Council (1997) furthermore aims to differentiate between different forms of hate speech, as far as intentions of the speaker are concerned, emphasizing the gravity of intentional use of hate speech for harmful purposes, such as publicly condoning or trivialising crimes of genocide, against humanity, and war crimes. The described legal framework proves that hate speech is a multifaceted phenomenon that is difficult to recognize, define and regulate.

Evolution of hate speech perception and its legal regulation

Today most of the (but not only) European countries regulate, limit or ban hate speech, thus reflecting a post-WW2 commitment of States to the protection of human rights, to the promotion of the right to personal dignity and freedom from discrimination, as enshrined in many international covenants and declarations.

However, such laws were developed long before 1945 and for very different reasons. They were in fact aimed at protecting governments and repressing political dissent, and were modelled after the English 1661 Seditious Act (that in turn originated in the thirteenth-century instrument of sedition), which imposed punishment against anyone who wrote, printed or preached words against the King. Different laws limiting hateful speech and banning expressions of incitement against the state or certain religious groups, expressions of disapproval or ridicule of the country's laws, expressions of discontent with the government, as well as defamation laws on honour insults existed in the late eighteenth century. The first modern hate speech codes were promulgated in France during the 1820s and 1830s, to oppose emerging socialist and workers movement. After the 1848–1849 uprisings in Germany, Prussia enacted a hate speech law in 1851, which later served as the basis for German penal code of 1871 (Goldberg, 2015, p. 482). It criminalized endangering the peace through the 'incitement to violence of different classes of population'. German criminal code (1998) criminalized hate speech against the state to enforce order, preserve public peace against political dissent, and to repress class, ethnic, national and political differences. Hate speech law was seen as a symbol of the struggle for power between the authority and the opposition.

Towards the end of the nineteenth century, the threat of antisemitism became tangible in Germany and to defend themselves, the Jewish activists gathered in an extra parliamentary defence and advocacy group. They started to promote a new interpretation of the article 130 by claiming that the Jewish minority represented a class deserving protection under article 130. By exposing antisemitic hate speech, they forced the law in a new direction, and ushered in a legal and political paradigm shift. Goldberg thus successfully shows that a new attitude to hate speech appeared already in the 1890s, which (contrary to the belief of most scholars, who argued the turning point happened no sooner than in the post-WW2 reaction to Nazism and Fascism) marks the first key turning point toward a human-rights model of hate speech law (Goldberg, 2015). The aim of hate speech legislation has changed through time, as has the understanding of hate speech itself. In hindsight we can say that hate speech evolved from expressions of discontent, defamation, dissent, and critique of the authority, to include all types of identity-based forms of discriminatory expressions. Changes in society have gradually moved the concept of hate speech regulation from repression towards the protection of human rights. The end of WW2 with emphasis on democracy and the dignity of human life represents an important milestone in this evolution.

The development of media and communication technologies has opened the doors to a parallel virtual universe, where proliferation of hate speech is substantially easier, faster, cheaper and more difficult to combat. Hateful messages disseminated online have become very common and call for adequate response and adoption of targeted measures that take into account the specific nature of digital communication. One of the challenges is the imperative to remove hateful content from the web as soon as possible, to limit the potential damage inflicted on the target of hateful comments and to prevent empowering

the perpetrators (Oboler in Gagliardone et al., 2015, p. 13). However, hate speech can be itinerant and can find expression elsewhere, even when the content is removed; its endurance is possible also because of the low cost, its potential for revival, and long-lasting relevance in particular spheres of discourse (Gagliardone et al., 2015, p. 14). Longevity of hateful comments online is better regulated in platforms such as Snapchat, where users' comments disappear after a while, thus providing a shorter time exposure and a more limited circle of influence. Due to the fact that the internet is not governed by a single entity, and platforms such as Facebook or Twitter are privately owned, the regulation of content on the internet presents a substantial challenge, in particular due to the unprecedented increase in the use of derogatory language on social media.

The language of discrimination and its reflections

A myriad of new challenges, such as migration crises, political upheavals, development of IT, all contribute to increasing social polarization and to an uninhibited use of hostile rhetoric. The language we are exposed to becomes subconsciously the language we use, the language we think in, and the language that constructs our reality. Language cannot be neutral – it reflects and structures our ideologies and worldviews. We respond to the images and language that surround us by recreating the same or similar images and patterns, using the same or similar language we have been exposed to. It is therefore safe to say that the influence of discriminatory language is not limited to the targeted group alone; it directly or indirectly affects all of us, as well as the entire social and belief system. Language is not merely an instrument for describing events, but a constituent part of events and can therefore influence political perceptions in a way that goes beyond its propositional content. Language is a tool of politics and is used not only to inform others about political issues, but also to persuade people in adopting specific courses of action in regard to these issues (Gorenc, 2009).

In its broad definition, derogatory language includes any type of insinuation and allegation about members of a given social category that explicitly ridicules or insults them. One of them is hate speech, which represents an extreme, expansive and fast-spreading form of language discrimination (Cervone et al., 2021, p. 81). A ground-breaking work on immigrant groups (Mullen & Smyth, 2004) reveals that derogatory labels are more numerous, more frequently used and more offensive, the smaller and the less familiar the social group is, and/or the lower its status in society. As observed by Stephan and Stephan (2000), these minorities may be seen as symbolic threats to cultural norms and values, even if they do not pose concrete threats over limited resources such as jobs. Discursive research has identified the shift of hate speech in far-right political discourse, which turned from demonizing people to demonizing belief systems. Verkuyten's (2013, p. 351) analysis of Geert Wilders' anti-Muslim rhetoric reveals his consistent rejection of hate accusations by stating he is against Islam as ideology, but has nothing against Muslims as people. Such discursive strategy allows far-right groups to project their identity as reasonable and rational, and their hostility as morally justifiable (Cervone et al., 2021, p. 83).

Social functions and consequences of derogatory language

Researchers have identified different psychological antecedents of derogatory language and hate speech such as right-wing authoritarian personality, social dominance

orientation, and the motivation to express prejudice (Forscher et al., 2015), as well as affective or emotional states specifically directed at the target: hate, contempt, disgust, and feelings of intergroup threat (Gerstenfeld, 2002). Moreover, five groups of likely social functions served by derogatory language have been isolated:

- (1) prejudice perpetuation against minority groups (Bianchi et al., 2019)
- (2) maintenance of status hierarchies (Rosette et al., 2013)
- (3) legitimization of violence against outgroup (Fasoli et al., 2015)
- (4) norm and role compliance (Carnaghi et al., 2011)
- (5) ingroup cohesion (Douglas, 2012).

Not all instances of derogatory language lead to physical violence; however, most forms of physical violence are preceded by verbal hostility. Furthermore, violence has different manifestations; physical, psychological, emotional, economic, and sexual violence are some of them. We cannot dismiss any as less harmful, just because it does not consist (at least initially) of physical abuse, even more so because the toll of verbal derogation can be seen in mental and physical health of victims. Derogatory language is also associated with risk behaviours, such as smoking, substance abuse and even attempted suicide. The research (Mullen & Smyth, 2004) reveals higher probability of suicide among immigrant groups targeted by ethnophaulisms. The increasing use of hate speech also coincides with increases in the reported incidence of hate crimes in the world.

The influence of hate speech does not extend only to targeted victims, but affects also bystanders, who can become desensitized and consequently perceive it as less offensive and more acceptable (Soral et al., 2018). Furthermore, derogatory language that bystanders are exposed to influences their own attitudes and behaviours, and causes their distancing from a targeted minority group and its members (Winieński et al., 2017). It is thus clear that hate speech not only affects the victims, but society in general. Politicians and other influential public persons abuse of their power and reach, to use hate speech for specific purposes. They actively contribute to the polarization of society, divert attention of public to the most convenient culprit (e.g. immigrants, homosexuals, protesters, artists, intellectuals ...), desensitize population and attack cohesion and general sense of community. Liberal democracies are supposed to give special protection to free speech; however, some types of speech can present a serious danger to social cohesion and democracy. This brings us to a dilemma whether to limit and regulate hate speech or allow for unrestricted freedom of expression, at the expense of potential discrimination of targeted groups.

Freedom of expression and its limitations

Freedom of opinion and expression is a fundamental human right, protected in Article 19 of the Universal Declaration of Human Rights (1948) and given legal force through all major international and regional human rights treaties.¹ International human rights law requires States to guarantee to all people the freedom to seek, receive or impart information or ideas of any kind, regardless of frontiers, through any media of a person's choice. The right to the freedom of expression is not an absolute right (ECHR, 2021),

and under exceptional circumstances (e.g. incitement to violence, hate speech and racism, Holocaust denial and references to Nazi ideology), the State may restrict the right under international human rights law (Bychawska, 2017); however, any restrictions on freedom of expression must be enshrined in law and precisely defined to serve a legitimate interest.

Hate speech triggers a confrontation of two important values, namely freedom of expression on one side, and the right of others to dignity and respect on the other. In the United States, the prevailing forces have mostly sided with unrestricted freedom of expression and against legal regulation of hate speech, referring to the First Amendment of the Constitution (1791). One of the most recognized advocates of the freedom of speech, John Stuart Mill (1978, p. 59) claimed that the fullest liberty of expression was necessary if we wanted to explore the true limits of our arguments, and not the limits of what is socially acceptable. Otherwise, the price we have to pay for this sort of intellectual pacification is the sacrifice of moral courage of the human mind. However, Mill also believed that limits on freedom of expression are justified, but only in the cases when preventing harm to others (1978, p. 36).

Countries define and understand freedom of expression in different ways, and cannot agree on when and how this freedom should be protected and when instead it should be limited (Wimmer, 2006). For a very long time, freedom of expression has been considered a precondition for a functioning and democratic society and has therefore enjoyed legal protection, with the exception of particular situations. Germany, for example, decided to limit the freedom of expression in the aftermath of the WW2, when censorship of pro-Nazi propaganda seemed the only moral and possible reaction. Laws banning the buying and selling of literature such as Hitler's *Mein Kampf*, as well as showing Nazi propaganda movies, e.g. *The Triumph of the Will*, were supported by many and passed in Parliament (Oppenheimer, 1998).

The complex balancing between the need to guarantee people the right to freely express themselves and to advocate their ideas on one side, and the defence of other people's right to be free from verbal abuse and to be protected as equal members of a society on the other, is not easy. Freedom of expression is the lifeblood of democracy; it facilitates open debate, a proper consideration of diverse interests and perspectives, and the negotiation and compromise necessary for consensual policy decisions. Efforts to suppress expression can allow unseen problems to fester and erupt in far more dangerous forms, which may lead to violence instead of ensuring peace and stability. The NGO Article19² believes that hate speech should be met at all times by counter claims, arguments and discussion. Suppressing it not only fails to resolve hatred but drives it underground and possibly encourages acts of violence (Coliver et al., 1992). When responding to hate speech, the States should not limit themselves to sanctioning, and should instead search for and target the underlying causes that drive hate speech, such as prejudice, intolerance, lack of information and similar. They should use positive measures to promote inclusive and intercultural dialogue, social plurality, respect and tolerance. Gagliardone et al. (2015) believe that counter speech constitutes a better way of blocking potential harm caused by hate speech. This is advocated also by Lepoutre (2017, p. 852), who admits that neither bans nor counter speech are without limitations and cannot be applied to all instances. What Lepoutre argues for is to rethink the use of bans in situations that might be better suited for counter speech. It seems that even

stringent legislation regarding hate speech focuses only on the extreme cases of its presence in society, leaving unpunished or unaddressed the cases of hateful expressions that do not constitute an offence under international law or those that do not qualify for criminal or civil sanctions, but are nevertheless offensive and problematic with a view to civility, tolerance and respect for others. It is therefore obvious that limiting hate speech with legal bans and restrictions does not suffice and we, as society, need to address it on all levels of social endeavour. Gagliardone et al. (2015) propose four types of initiatives to counteract the emergence and dissemination of hateful messages: monitoring and discussing hate speech; mobilizing civil society; lobbying private companies; media and information literacy campaigns.

Hate speech should be banned because it is inconsistent with the fundamental values of liberal democracy, such as equal respect for all people. Waldron (2010) disagrees with the view that the targets of hate speech should learn to put up with it, and that the freedom of speech is more important than minimizing the feelings of anxiety or distress of the targets. Hate speech namely possesses harmful tendencies that endanger social cohesion and injure the dignity of targeted groups. It should be banned due to the general apprehension of its effects and not only when there is evidence of substantial harm, caused to social order or its victims (Waldron 2010, p. 1650). This reasoning leaves a lot of (too much) room for manoeuvre to governments and legislators who decide whether an instance of hate speech requires legal intervention, as can be seen in various legal practices and court proceedings in different countries. One of the claims underpinning the idea that hate speech does not need special, additional regulation argues that instances of harmful and harm causing speech are already regulated and banned in various other legal instruments (Boonin in Van Mill, 2021).

Case study

Young people are increasingly becoming targets of different forms of cyberviolence, and with unprecedented surge in the use of digital technologies due to the Covid 19 restrictions, the negative consequences of cyber exposure have multiplied. The empirical part of the article draws from the answers of Slovenian social sciences students participating in an online survey on hate speech. Our goal was to find out how young people perceive and understand hate speech and its manifestations and how (if) they react to it. We were interested in revealing the channels of hate speech dissemination, the role and accountability of opinion leaders as hate speech perpetrators, as well as possible solutions. We examined how familiar the respondents were with programs and platforms addressing hate speech, and how efficient they believe non legislative measures against hate speech can be.

There were 499 respondents in total, with 309 (62%) of them completing the survey, which consisted of 26 questions. It started with four definitions of hate speech and respondents could choose more than one:

- (1) Expression of opinions and ideas that are inherently discriminatory and generally directed against disadvantaged groups (ethnic, national, religious, cultural, sexual, etc.) or against individuals who, because of their personal circumstances, belong to such groups. Hate speech dehumanizes those against whom it is directed and may also be intended to humiliate, intimidate or incite violence.

- (2) Promotion of hatred and violence against a certain group or its members based on their race, nationality, ethnicity, sexual orientation.
- (3) Active incitement to violence against a certain group or its members based on their race, nationality, ethnicity, sexual orientation.
- (4) Public promotion of hatred, violence and intolerance against certain social groups.

Definition (a) was selected by 81% respondents, (b) by 64%, (c) by 55% and (d) by 63% respondents. Only 55% understood definition (c) as hate speech, albeit active incitement implies a strong mobilization against someone.

Respondents were then asked to evaluate six statements published in Slovenian media and included in the survey anonymously. They had to specify whether a statement presented or not an instance of hate speech. If they answered 'yes', they were asked for additional explanation: the statement (a) incites hatred of particular social group, (b) incites violence to a particular social group, (c) encourages intolerance to a particular social group). If they answered 'no', they were also asked to explain the reason for their decision: the statement (a) expresses personal opinion we are all entitled to, (b) is an example of inadequate expression, however, it is not contentious substance wise, (c) is simply true. Targets of the six statements were independent cultural producers, immigrants, 'the erased',³ the Catholic church and LGBTQ community. The statements were as follows:

- (1) These are unemployed people living at taxpayers' expense, especially self-proclaimed cultural producers who receive unjustified financial support instead of working.
- (2) Immigrants violate legislation and illegally cross the borders. They attack the rule of law and sovereignty of the country they are entering.
- (3) LGBTQ identity is not natural and should be diagnosed as a disease.
- (4) All the erased are invited to a ball, where Jelinčič⁴ will play the machine gun.
- (5) There are things one simply should hate. I believe Catholic church is something you have to hate. I feel it is my duty as a citizen.
- (6) Approach allowed max. 500 m to the border. Whoever comes closer will be shot, God will recognize his own.

Most respondents identified hate speech in statements inciting to violence (no.4, 70% and no. 6, 83%; 40% of the former and 69% of the latter believe statements promoted violence against a certain social group). Another statement identified as hate speech by 76% of respondents targeted LGBTQ community; 42% believed the statement promoted hatred to a specific social group and 32% that it promoted intolerance to a specific social group. This shows that despite negative propaganda by authorities, the claim that LGBTQ identity is a disease, sounds too anachronistic and discriminatory. Answers to statement no. 2 show some inconsistency in the respondents' attitude to immigrants, as only 56% answered the statement is an instance of hate speech and openly discriminates a specific social group. It seems that intolerance is not perceived as bad enough to activate feelings of solidarity in respondents. Statement no. 5 was labelled as hate speech by 69%; the majority of them believed the statement incited hatred against a particular social group. The Catholic church was thus identified as a targeted social group. The most divisive statement (no. 1) targeted independent cultural producers. Forty-six

percent of respondents labelled it as hate speech and 48% believed it could not be classified as such. It is important to say that for the past two years independent culture producers have been constantly under attack of government and its supporters, and similar statements have become a standard in political discourse. This could explain the perception of respondents, for it has been proven that hate speech also exerts influence on bystanders who become accustomed to it (Soral et al., 2018) and actually dissociate themselves from the targeted group (Winiewski et al., 2017). It was interesting to see the perception of respondents regarding the presence of hate speech in media. Thirty-six percent believed the share of hate speech in social and traditional media is similar, 50% believed there is more hate speech in social media, and 7% believed there is less hate speech in social media. Eighty-five percent of respondents believed the level of hate speech in social media would decrease, were it mandatory to sign comments with full name. Sixty-two percent of respondents also believed that opinion leaders and politicians set an example with their behaviour and significantly influence public perception of what is or is not acceptable. Eighty-seven percent of respondents believed that unsanctioned use of hate speech by opinion leaders and politicians is reflected in hateful public discourse, and 89% believed that politicians and opinion leaders should use a more respectful discourse. Sixty-four percent of respondents supported the idea that hate speech should be prosecuted and believe this would not represent a form of censorship.

The last part of the survey examined how familiar the respondents were with the activities of civil society and NGOs in the field of hate speech regulation. The hotline *Spletno oko* ('WebEye') is a part of Safer Internet Program, primarily focused on the prevention of child abuse, which includes also other forms of violent behaviour and expression, such as hate speech. People can report instances of hate speech to *Spletno oko*, and if the reported instance presumably contains illegal elements, it is referred to police. Compared to 2019, *Spletno oko* reported a threefold increase in cases of illegal hate speech. Most of them alluded to other personal circumstances, such as political background. There was about ¼ of xenophobic speech mostly targeting immigrants, and about 13% was hate speech on a religious basis. In most cases the hateful expression took an extreme form by promoting violence, inciting to murder, or extermination of an entire social group (or its members). In 33% of cases hate speech included denial of holocaust and glorification of war crimes. Sixty-three percent of reported contents were published on news portals, a third on social media, and the rest on internet forums (Valentič & Motl, 2020). Another initiative aimed at exposing and thus curbing the dissemination of hateful content was a grassroots initiative, the webpage *ZLOvenia* (EVILSlovenia), where instances of hate speech were exposed. It was particularly active at the time of migration crisis in 2014, when extremely hateful expressions flooded the web. Authors of hate speech were publicly exposed with their full name and photo on the webpage, as well as on printed and affixed posters on the walls or billboards. The page was supposed to close down when things changed for the better, but actually stopped existing already in 2016. Contrary to the common belief that hate speech perpetrators disregard the reaction of public and engage in the dissemination of hateful content despite it, their exposure on the webpage and on posters convinced many to contact the webpage, express regret and ask for the removal of their hateful post together with personal information. However, survey results regarding *Spletno oko* and *ZLOvenija* are pessimistic. If the lack of information about the webpage *ZLOvenija*, which ceased to exist in 2016, can be

explained by the young age of our respondents at that time, we cannot overlook the fact that only 18% of the respondents knew about Spletno oko. Judging by the survey response, it is not easy to recognize and implement non-legislative measures aiming at hate speech regulation.

The last group of questions referred to active response (or the lack of it) of survey participants to hate speech. Only 38% have already reacted to hate speech (1% reported it to Spletno oko, 1% to police, 8% to the editor of webpage or media outlet, and 28% responded directly to the author of hateful comment or post). 54% of respondents have never decided to react, as they believe that: nothing would change (35%), the editor or police would not react (13%), they could be in danger if reporting (4%), hate speech does not represent a problem (2%).

Based on the survey results we can say the proposals of Gagliardone et al. (2015) should be seriously considered, addressed and included in the agenda of politicians, educators, media, civil society, and NGOs, which must all activate and strengthen their response to hate speech. Media and information campaigns must be promoted in educational institutions of all levels, while children and youth should be taught how to recognize and react to hateful messages, and why it is wrong to perpetrate or condone hate speech. Furthermore, hate speech perpetrated by politicians should not be tolerated and should be strictly sanctioned.

Conclusions

Already the first institutionalized forms of communication, e.g. religion, have limited the right of individuals to freely express their opinion regarding the matters of authority and the ruling elites (Splichal, 1997, p. 351), and for a long time the focus of debates on the unlimited freedom of expression, regardless of the content, reflected this history of repression. However, with time and especially in response to the development of media and communication technologies, 'freedom of expression' started to show its uglier face, with the aim of discriminating against and offending particular social groups, its members, or individuals. Due to easy access, low cost and anonymous participation, social media have become the leading platform for public debate. From inclusive and seemingly democratic beginnings, they have evolved into platforms and algorithms that follow our online presence, analyse our interests, push targeted advertisements, and help create echo-chambers of like-thinking actors. They drive social polarization to its extreme and represent a fast and efficient highway for the dissemination of hate speech and brutalization of public discourse. The more hateful public discourse people are exposed to, the more accustomed and less empathic they become. This can explain low reaction rates to hate speech and the lack of empathy for the targeted groups revealed also by the results of the conducted survey. Respondents have identified the harmful consequences of hateful discourse in which Slovenian politicians and public persons regularly engage. They supported the claim that politicians abuse their influence to spread hateful messages without inhibition and set an example to the public, which reacts by reproducing the same discourse and by distancing itself from the victims of hate speech and not from the perpetrators.

In recent years the world has been faced with serious challenges that strengthen the sense of insecurity and further polarize societies. 'Post truth' reality makes it increasingly

difficult for people to distinguish truth from lies, and hate speech rides the wave of manipulation and arbitrary interpretation of truth. In theory, mature democracies should be able to address the problem of hate speech with measures other than legislation and criminalization, namely by educating and informing people about hate speech, its manifestations and its causes. However, to bridge the gap between theory and reality, the states should implement positive measures conducive to tolerant and inclusive dialogue and behaviour. In doing so they need the support of an empowered civil society, able and willing to act upon all social manifestations of hate and intolerance by exposing haters as an extreme and despicable minority and by promoting the values of a democratic, pluralistic and respectful society. We are civil society.

Notes

1. See Article 19 of the International Covenant on Civil and Political Rights (ICCPR); Article 9 of the African (Banjul) Charter on Human and Peoples' Rights (ACHPR); Article 13 of the American Convention on Human Rights (AmCHR), and Article 10 of the European Convention on Human Rights (ECHR).
2. ARTICLE 19 is an international human rights NGO promoting freedom of expression and freedom of information worldwide.
3. The erasure from the register of permanent residents implemented by the administrative bodies of the Republic of Slovenia was an arbitrary act that did not have any basis in the law, as was established by the Constitutional Court. The erasure mainly (but not exclusively) affected people born in other republics of the former Yugoslavia who had Yugoslav citizenship and also citizenship of another republic of the former Yugoslavia, but lived in the former Socialist Republic of Slovenia where they had permanent addresses. When Slovenia became an independent country, on 25 June 1991, citizens of the former Socialist Republic of Slovenia automatically became citizens of the new country, the Republic of Slovenia. Further, according to Article 40 of the Citizenship of the Republic of Slovenia Act, all citizens of other republics of the former SFRY with permanent addresses in the Socialist Republic of Slovenia had the right to apply for Slovenian citizenship within six months from the date of independence. Those who did not obtain citizenship (because they failed to apply for whatever reason, or their application was refused or discarded, or the procedure was terminated), lost their permanent residence status. They were deprived of their permanent residence status permanent resident status: <https://www.mirovni-institut.si/izbrisani/en/about-erasure/index.html>
4. Zmago Jelinčič Plemeniti is a Slovenian politician, founder and president of Slovenian National Party (March 1991), which won more than 11% of votes in the first elections to the Slovenian National Assembly in 1992, largely due to his hateful rhetoric against people from former Yugoslavia.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Nina Gorenc is a Lector of English and Italian language, and a PhD in Political Science. Her interests lie in interdisciplinary manifestations and intersections of language, political science and communication, which is also reflected in her academic career. Her research interests span from American presidency, political rhetoric, cultural diplomacy, to discourse & content analysis, post truth society and hate speech. She has served as the chair of Language department at the

Faculty of Social Science since 2015, where she is also employed. Her professional, research and educational activities are divided between Slovenia and Italy, where she regularly participates in research and educational projects, or as a visiting scholar. She lives with her family in Ljubljana, a perfect destination for sports and outdoor activities that she loves.

ORCID

Nina Gorenc  <http://orcid.org/0000-0003-1872-5547>

References

- Bayer, P., & Bard, P. (2020). *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*. European Parliament. <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/655135>
- Bianchi, M., Carnaghi, A., Piccoli, V., Stragà, M., & Zotti, D. (2019). On the descriptive and expressive function of derogatory group labels: An experimental test. *Journal of Language and Social Psychology*, 38(5-6), 756–772. <https://doi.org/10.1177/0261927X19867739>
- Bychawska, S. D. (2017). *Protecting the right to freedom of expression under the European convention on human rights: A handbook for legal practitioners*. Council of Europe. <https://rm.coe.int/handbook-freedom-of-expression-eng/1680732814>
- Carnaghi, A., Maass, A., & Fasoli, F. (2011). Enhancing masculinity by slandering homosexuals: The role of homophobic epithets in heterosexual gender identity. *Personality and Social Psychology Bulletin*, 37(12), 1655–1665. <https://doi.org/10.1177/0146167211424167>
- Cervone, C., Augoustinos, M., & Maass, A. (2021). The language of derogation and hate: Functions, consequences, and reappropriation. *Journal of Language and Social Psychology*, 40(1), 80–101. <https://doi.org/10.1177/0261927X20967394>
- Coliver, S., D'Souza, F., & Boyle, K. (1992). *Striking a balance: Hate speech, freedom of expression and non-discrimination*. University of Essex, Human Rights Centre.
- Constitution of the United States, Amendment I. (1791, December 15). <https://constitution.congress.gov/constitution/amendment-1/>
- Council of Europe, Committee of Ministers . (1997). Recommendation No. R (97) 20 of the Committee of Ministers to members states on “hate speech” adopted on 30.10.1997.
- Douglas, K. M. (2012). *Psychology, discrimination and hate groups on-line*. *Oxford Handbook of Internet Psychology*. Part of <https://doi.org/10.1093/oxfordhb/9780199561803.013.0011>
- European Commission against Racism and Intolerance. (2015). General Policy Recommendation no. 15 on combating hate speech. <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01>
- European Convention on Human Rights. (2021). Council of Europe & European Court of Human Rights. https://www.echr.coe.int/documents/convention_eng.pdf
- Fasoli, F., Carnaghi, A., & Paladino, M. (2015). Social acceptability of sexist derogatory and sexist objectifying slurs across contexts. *Language Science*, 52, 98–107. <https://doi.org/10.1016/j.langsci.2015.03.003>
- Forscher, P. S., Cox, W. T., Graetz, N., & Devine, P. G. (2015). The motivation to express prejudice. *Journal of Personality and Social Psychology*, 109(5), 791–812. <https://doi.org/10.1037/pspi0000030>
- Framework decision on combatting racism and xenophobia. (2008). *OJ L* 328, 6.12.2008.
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000233231>
- German Criminal Code. (1998). Federal Law Gazette I, p. 3322. https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html
- Gerstenfeld, P. (2002). A time to hate: Situational antecedents of intergroup bias. *Analyses of Social Issues and Public Policy*, 2, 61–67. <https://doi.org/10.1111/j.1530-2415.2002.00027.x>

- Goldberg, A. (2015). Hate speech and identity politics in Germany, 1848–1914. *Central European History*, 48(4), 480–497. <https://doi.org/10.1017/S0008938915000886>; <http://www.jstor.org/stable/43965202>
- Gorenc, N. (2009). *Presidential justifications for military interventions abroad* [Doctoral dissertation]. University of Ljubljana. Repository of the University of Ljubljana. http://dk.fdv.uni-lj.si/doktorska_dela/pdfs/dr_gorenc-nina.PDF
- Lepoutre, M. (2017). Hate speech in public discourse: A pessimistic defense of counterspeech. *Social Theory and Practice*, 43(4), 851–883. <http://www.jstor.org/stable/26405309> <https://doi.org/10.5840/soctheorpract201711125>
- Mill, J. S. (1978). *On liberty* (Kindle ed.). Hackett Publishing Company, Inc. <https://www.amazon.com/Liberty-John-Stuart-Mill-ebook/dp/B004UJ8LVM>
- Mullen, B., & Smyth, J. M. (2004). Immigrant suicide rates as a function of ethnophobias: Hate speech predicts death. *Psychosomatic Medicine*, 66(3), 343–348. <https://doi.org/10.1097/01.psy.0000126197.59447.b3>
- Oppenheimer, P. (1998). In the name of democracy. *Index on Censorship*, 27(1), 52–56. <https://doi.org/10.1080/03064229808536292>
- Rosette, A. S., Carton, A. M., Bowes-Sperry, L., & Hewlin, P. F. (2013). Why do racial slurs remain prevalent in the workplace? Integrating theory on intergroup behavior. *Organization Science*, 24(5), 1402–1421. <https://doi.org/10.1287/orsc.1120.0809>; <http://www.jstor.org/stable/42002912>
- Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136–146. <https://doi.org/10.1002/ab.21737>
- Splichal, S. (1997). *Javno mnenje: teoretski razvoj in spori v 20. stoletju* (1. izd.). Fakulteta za družbene vede.
- Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Lawrence Erlbaum Associates Publishers.
- Treaty on the Functioning of the European Union. (2012). *OJ C 326*, 26.10.2012, pp. 47–390.
- United Nations. (1948). *Universal declaration of human rights*. <https://www.un.org/en/about-us/udhr/foundation-of-international-human-rights-law>
- United Nations. (1964). International convention on the elimination of all forms of racial discrimination. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-elimination-all-forms-racial>
- United Nations (General Assembly). (1966). International covenant on civil and political rights. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>
- Valentič, U., & Motl, A. (2020). *Letno Poročilo Spletno oko 2020*. Center za varnejši Internet Prijavna Točka Spletno Oko. https://www.spletno-oko.si/sites/default/files/letno_porocilo_spletno_oko_2020.pdf
- Van Mill, D. (2021). Freedom of speech. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2021 ed.).
- Verkuyten, M. (2013). Justifying discrimination against Muslim immigrants: Out-group ideology and the five-step social identity model. *British Journal of Social Psychology*, 52(2), 345–360. <https://doi.org/10.1111/j.2044-8309.2011.02081.x>
- Waldron, J. (2010). Dignity and defamation: The visibility of hate. *Harvard Law Review*, 123(7), 1596–1657. <http://www.jstor.org/stable/40648494>
- Wimmer, K. (2006). Toward a world rule of law: Freedom of expression. *The ANNALS of the American Academy of Political and Social Science*, 603(1), 202–216. <https://doi.org/10.1177/0002716205282120>
- Winiewski, M., Hansen, K., Bilewicz, M., Soral, W., Swiderska, A., & Bulska, D. (2017). *Contempt speech, hate speech: Report from research on verbal violence against minority groups*. Stefan Batory Foundation.