

Paz-Rebollo, María Antonia & Segado.Boj, Francisco (2025), Hate Speech. In Nai, A., Grömping, M., & Wirz, D. (Eds). *Elgar Encyclopedia of Political Communication*. Edward Elgar Publishing. Accepted version.

Hate Speech

María Antonia Paz-Rebollo, Complutense University, mapazreb@ucm.es

Francisco Segado-Boj, Complutense University, fsegado@ucm.es

The rise of online hate speech has been fuelled in several countries by political polarisation and the growth of the alt-right movement. While this form of hate speech is not a criminal offence, it has serious psychological and social consequences for vulnerable target groups. There is no universally accepted definition of hate speech, but some indicators can be identified, including sexist language, denial of human rights, incitement to violence and negative stereotyping of minorities. Hate speech relies on simplistic explanations and discursive strategies that promote group inclusion/exclusion and antagonism. They are rooted in contemporary populist movements and can lead to stochastic terrorism. These can also affect public opinion, shift the Overton Window of acceptable Political Possibility, and influence the media agenda. While authorities, journalists, cultural actors, and citizens are all involved in combating hate speech, this raises a debate about freedom of expression.

Hate speech; Social media; Narratives; Stochastic terrorism; The Overton Window; Control and moderation.

This entry was made within the framework of the research project PID2019-105613GB-C31, “Cartografía de los Discursos de Odio en España desde la Comunicacion, funded by Ministerio de Ciencia e Innovación (Spain).

Despite the various control measures in place to restrict it, the presence of online hate speech has increased in recent decades. This has been due to the ease with which it can be disseminated and to the fact that online hate speech is encouraged by political polarisation and the growth of the alt-right movement in different countries. Even though hate speech might remain associated mainly to extreme, partisan groups, some of their topics, narratives and strategies have been adopted by mainstream political actors and thus become a common feature in contemporary political communication (Said-Hung, Moreno-López & Mottareale-Calvanese, 2023).

We will first give an overview about definitions on hate speech and then, we will analyse the narratives and rhetoric resources as well as the ideological content on which hate speech is built. Next, we will summarize how hate speech influence's public opinion. Last, we will introduce the main strategies to fight hate speech on social media. There is no universally accepted definition of hate speech. This concept is constantly being reworked in public and political debates because of its complex nature.

Papcunová et al. (2021) proposed a number of indicators for identifying it, including the use of sexist language, the denial of human rights, incitement to violence, attribution of negative stereotypes to a minority, or use of insulting labels to refer to a group on the basis of race, national origin, religious affiliation, gender or sexual identity, among others.

Hate speech may be formulated with different objectives in mind: expressing an opinion, generating debate, harming a person or group, or seeking support for a cause. They also have different intensities. The Anti-Defamation League proposes five levels in the form of a pyramid, distributed from the lowest to the highest intensity. At the base of the pyramid of hate are negative stereotypes ('Women should stay in the kitchen'), followed by insults ('Gypsies, disgusting sub-humans') and discriminatory expressions ('Why have a child if you can't afford to care for one'). At the top of the list are threats ('a good beating and fuck off to where you came from') and genocide ('What we should do is gas the 3,000 [a slum area] and kill all the insects there').

Narratives, Ideology and Hate Speech

Hate speech is usually an impulsive and emotional type of discourse. It conveys antagonistic, negative feelings and attacks people for who they are, often converging as intersecting forms of hatred. For example, Islamophobia amalgamates denigrating expressions that refer to religion, ethnicity, politics and even gender. Subtle means are used for legitimization purposes and to circumvent moderation controls. For example, humour, usually accompanied by emoticons, disqualifies, humiliates, or even proposes violent actions, but the smile it elicits can minimise the perception of how serious it is. Also noteworthy is the use of dehumanising metaphors (Musolff, 2015) that identify certain groups (migrants or people living in poverty, for example) with small, insignificant animals that are easy-to- exterminate pests ('cockroaches' or 'rats'). Finally, it is worth mentioning the presence of images, specifically memes, which are a

powerful and versatile tool in political persuasion and propaganda campaigns (Paz-Rebollo, Mayagoitia-Soria and González-Aguilar (2021).

These discourses rely on simplistic explanations and discursive strategies based on dichotomies in which there is always a positive self-representation, while the other or others are not only presented negatively, but demonised. Hate speech is produced in terms of group inclusion/exclusion, fostering antagonism. This is how xenophobia rears its head, and also how hostility towards women is expressed in the manosphere (Lacalle, Gómez-Morales, Vicent-Ibáñez, 2023) and how differences based on religious belief or sexual identity are targeted, among others.

Hate speech is deeply rooted in contemporary populist movements. This symbiotic relationship between hate speech and populism is defined by the conception of a confrontation between the real and 'true' people (us) and their enemies (them) (Casero-Ripollés, Sintes-Olivella & Franch, 2017). Depending on the political orientation of the populist movement in question, these enemies can be vague entities (such as 'the ruling elite', 'traditional politicians' or 'Brussels bureaucrats'), opposing ideological movements (the 'Woke' movement, 'libtards'), or as is often the case in right-wing populist movements, minorities such as immigrants or LGBTQ people. As they attack these enemies, populists try to rally support around their identity by delegitimising and even dehumanising other members of society.

According to populist logic, the 'real people' can face two different types of enemies: the internal enemy and the external enemy (Bjånesøy, IVarfalten & Berntzen, 2023). The internal enemy is a group of people who, more or less formally, are part of society, but whose interests or behaviour are detrimental to the 'real' interests and benefits of the 'true people'. In these cases, hate speech is often directed against traditional politicians, opposition political parties or sympathisers of certain ideologies. For example, feminists may be labelled 'crazy', 'ugly' or 'authoritarian', or government parties may be identified as 'parasites'. LGBTQ+ people are often qualified as internal enemies, portrayed as 'degenerates' and accused of destroying 'traditional values' or 'the traditional family'.

The target groups considered 'external enemies' are mainly not formally part of society but are very often immigrants and foreigners. They are depicted as outsiders, unable to fit into the country, usually displaying anti-social or illegal behaviour and, once again, as destroying the fundamental values and identity of the nation.

Public opinion and hate speech

Hate speech can affect public opinion in a number of ways. Among these, the most directly impactful and seriously harmful is the promotion of so-called stochastic terrorism (Amman & Meloy, 2022). This phenomenon refers to the indirect encouragement to the commission or instigation of attacks against certain groups or individuals by subjects who are not necessarily part of a terrorist organisation or formally registered with an extremist party.

The dissemination and consumption of messages and discourses that dehumanise minorities and portray them as enemies of the majority or as ‘degenerates’ prompt the idea among some members of the public that someone has to act against these groups. Action may be taken in the most extreme case, or among the most unbalanced members of such collectives. Thus, hate speech either online or in the media can lead to offline physical violent hate crimes, similar to terrorist attacks. This effect has been observed in members of extreme and marginalised online communities but may also arise from the constant stream of hate speech disseminated in the mainstream, or at least niche, media (Munn, 2019).

One of the best-known cases of stochastic terrorism is the 2019 shooting at the Christchurch Mosque in New Zealand, where 51 people were killed. The killer was an active member of far-right forums and a believer in the ‘Great Replacement’ conspiracy theory, which equates immigration with an invasion, as part of a movement to eliminate the white population and replace it with other ethnicities. His involvement in the consumption (and production) of hate speech explains, at least partially, his involvement in and motivation for the killing.

However, hate speech circulating in less marginal and more mainstream environments can also trigger such acts of stochastic terrorism. This was the case of the Club Q shooting. In 2022, a young man shot five people dead in a gay club in Colorado Springs (USA). This act has been accused of being inspired by a stream of speeches and opinion pieces in the far-right US partisan media that claimed that the homosexual community had engaged in paedophilic practices.

The effects of hate speech can also have more indirect, less violent consequences, but also with disturbing implications. In this sense, the presence of hate speech messages (especially in the mass media) can displace the so-called ‘Overton window of political possibility’ (Russell, 2006). According to this theory, while a potentially unlimited

number of measures and policies can be proposed and discussed, a society, at any given moment in history, can only accept a limited set of such propositions. The ‘Overton window’ covers the set of political ideas that could be debated in the public sphere, and those deemed unacceptable would be outside the window. However, the ‘Overton window’ is not fixed; it can be widened or narrowed and can be shifted to include or exclude some options or proposals from the policy arena.

The presence of hate speech against certain groups can therefore displace this ‘Overton window’ to reintroduce options that limit or restrict civil rights or liberties into the political debate as reasonable. For example, countries that had accepted same-sex marriage, where it was accepted that this right could not be taken away, reintroduced the debate or even repealed those laws thanks to the displacement of the Overton Window. Minorities may thus be left further unprotected and mistreated (Ikeanyibe et al., 2018). Along similar lines, hate speech can also influence the third-level agenda setting (Su & Yu, 2020). This theory postulates that the media can influence the issues that are debated in the public sphere. Broadly speaking, agenda setting theory explains that the media can influence public opinion by introducing issues into the public debate, as the media focus is what the public focuses on. They do so by attributing characteristics to issues or actors (for example, by portraying a political candidate as honest or dishonest) and by linking or connecting issues together. For example, the media may construct and promote the notion that immigration is linked to crime, even if this connection is not explicitly mentioned in a news story. However, by presenting two or more issues together, the media create a topical connection about them that is conveyed to the public.

Fighting Hate Speech in social media

The responsibility to combat hate speech lies with the authorities, journalists and cultural and social actors, as well as with citizens. While some initiatives have been taken at the international level to fight this phenomenon (Commission Recommendation (EU) 2018/334, the EU Code of Conduct on Countering Illegal Hate Speech Online, among others) with positive results, they have raised an intense debate about freedom of expression, also present in the academic literature on hate speech (Tontodimamma et al., 2021). The American model upholds freedom of expression and only censors’ speech that directly leads to the commission of a crime; and the European model

condemns hate speech on the grounds that any discourse against the equality and dignity of minority groups is a crime, even if there is no direct incitement to violence.

For their part, social networks and the digital press have issued specific rules to prevent uncivil or hateful comments and monitor their compliance through algorithms that detect especially foul language and insults. In addition to automatic detection, there are specialised companies and/or community managers who manually ban inappropriate messages (authoritarian moderation); or interact online to redirect and improve those comments and encourage positive attitudes (engagement moderation). The obligation to pay to read and comment (subscription to the press) and to provide personal data (loss of anonymity) is considered decisive in the decline of these expressions but does not lead to their complete disappearance. Finally, it is hoped that users will co-operate by being alert to the existence of these expressions and responding with counter-discourse or direct confrontation. However, while 'good citizens' are more likely to oppose online hate (Kunst et al., 2021), they often choose to stay out of such situations because of the social costs of confrontation. Automated hate detection techniques are not always effective either because users create new words (by adding signs, spaces, numbers) and because moderators have a different appreciation of online hate speech. Even some forms of incivility are accepted by journalists because they are related to polarisation and are therefore newsworthy.

In short, hate speech is a major public concern, as it hinders the high-quality, respectful debate that is characteristic of democratic societies. It has permeated the current civic and political discourse and has become a common feature on political campaigning. Researchers need thus to focus on the different effects of this phenomenon on public opinion as well as on strategies to mitigate it.

References

- Amman, M., & Meloy, R. (2022). Incitement to Violence and Stochastic Terrorism: Legal, Academic, and Practical Parameters for Researchers and Investigators. *Terrorism and Political Violence*, 36(2), 234-245.
<https://doi.org/10.1080/09546553.2022.2143352>
- Bjånesøy, L., Ivarsflaten, E., & Berntzen, L. E. (2023). Public political tolerance of the far right in contemporary Western Europe. *West European Politics*, 46(7), 1264-1287. <https://doi.org/10.1080/01402382.2023.2167046>

- Casero-Ripollés, A., Sintes-Olivella, M., & Franch, P. (2017). The populist political communication style in action: Podemos's issues and functions on Twitter during the 2016 Spanish general election. *American Behavioral Scientist*, 61(9), 986-1001. <https://doi.org/10.1177/0002764217707624>
- Ikeanyibe, O. M., Ezeibe, C. C., Mbah, P. O., & Nwangwu, C. (2018). Political campaign and democratisation: Interrogating the use of hate speech in the 2011 and 2015 general elections in Nigeria. *Journal of Language and Politics*, 17(1), 92-117.
- Lacalle, C.; Gómez-Morales, B., Vicent-Ibáñez, M. (2023): Misogyny and the construction of toxic masculinity in the Spanish Manosphere (Burbuja.info) *Profesional de la Información*, 32(2). <https://doi.org/10.3145/epi.2023.mar.15>
- Kunst, M., Porten-Cheé, P., Emmer, M., Eilders, C. (2021). "Do 'Good Citizens' fight hate speech online? Effects of solidarity citizenship norms on user responses to hate comments". *Journal of Information Technology & Politics*, 18(3), 258–273. <https://doi.org/10.1080/19331681.2020.1871149>
- Munn, L. (2019). Alt-right pipeline: Individual journeys to extremism online. *First Monday*, 24(6). <https://doi.org/10.5210/fm.v24i6.10108>
- Musolff, A. (2015). Deshumanizing metaphors in UK immigrant debates in press and online media. *Journal of Language Aggression and Conflict*, 3(1), 41-56. <https://doi.org/10.1075/jlac.3.1.02mus>
- Papcunová, J., Martončík, M., Fedáková, D., et al. (2021). Hate speech operationalization: a preliminary examination of hate speech indicators and their structure. *Complex & Intelligent Systems*, 9(3), 91-16. <https://doi.org/10.1007/s40747-021-00561-0>
- Paz-Rebollo, María Antonia; Mayagoitia-Soria, Ana; González-Aguilar, Juan Manuel (2021). "From Polarization to Hate: Portrait of the Spanish Political Meme". *Social Media + Society*, 7(4). <https://doi.org/10.1177/20563051211062920>
- Tontodimamma, A., Nissi, E., Sarra, A., & Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126, 157–179. <https://doi.org/10.1007/s11192-020-03737-6>
- Russell, N. J. (2006, January 4). An introduction to the Overton window of political possibilities. *Mackinac Center for Public Policy*. <https://www.mackinac.org/7504>

- Said-Hung, E., Moreno-López, R., & Mottareale-Calvanese, D. (2023). Promotion of hate speech by Spanish political actors on Twitter. *Policy & Internet*, 15(4), 665-686. <https://doi.org/10.1002/poi3.353>
- 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>
- Su, Y., & Hu, J. (2020). A territorial dispute or an agenda war? A cross-national investigation of the network agenda-setting (NAS) model. *Journal of Information Technology & Politics*, 17(4), 357-375. <https://doi.org/10.1080/19331681.2020.1756553>