

# Hate Speech in the Time of COVID-19

Kathryn Benier, Matteo Bonotti, and Joe Collings-Hall



## SUMMARY

This brief provides an overview of the academic literature on hate speech in order to offer analytical tools for critically assessing the rise in hate speech incidents during the current COVID-19 pandemic. We focus, first, on how hate speech is generally defined and understood, showing that there is no consensus in this area. We then examine arguments for and against the regulation of hate speech, before moving on to analyse the key triggers and causes of hate speech. We focus especially on crises which, like the current COVID-19 pandemic, may provide fertile ground for the increase in hate speech incidents.

## Introduction

Alongside public health and economic costs, the ongoing COVID-19 crisis has also resulted in a broad range of social harms. Prominent, among the latter, has been a sharp increase in incidents of hate speech targeting members of various ethnic and religious minorities. While people of Chinese and East Asian origins have often been the main targets of hate speech during the current pandemic,<sup>1</sup> other targeted minorities include Muslims<sup>2</sup> and Jews.<sup>3</sup> In this research brief we provide a framework for understanding these phenomena by examining what hate speech is and how it is conceived in different jurisdictions; whether and why it should be regulated; and what causes it, especially in times of crisis.

## Definitions and Legislation

Hate speech has been described in various ways, for example as involving basic disrespect and as an attack on its victims' dignity and human rights; as inciting violence and other criminal acts, thus also often undermining public peace; as producing a climate of hatred; as communicating insulting, abusive and/or disdainful messages; as humiliating, subordinating and/or stigmatising its victims; and as reinforcing discrimination, marginalisation and other injustices already endured by members of various vulnerable groups (Brown 2015; Strossen 2018). These diverse understandings of hate speech are also reflected in the variety of laws and regulations across different jurisdictions, resulting in the 'fragmentation into separate, context-oriented, historically biased, culturally defined, politics-shaped, country-specific approaches to speech restrictions' (Haraszti, 2012, p. xiv).

In some countries, hate speech is characterised as a denial of human rights. For example, hate speech may infringe upon the right to non-discrimination, as when a 'Whites Only' sign is displayed on a restroom door in a workplace (*EEOC v. Tyson Foods, Inc.* 2006). Other jurisdictions target so-called 'expression-oriented hate speech' such as 'fighting words' (i.e. words that cause an immediate, often violent, reaction by the listener) (*Chaplinsky v. New Hampshire* 1942) or cross-burning, which is historically connected with racism and violence against African-Americans in the US (*Virginia v. Black* 2003).

Other hate speech laws and regulations target stereotyping or stigmatising others. Some countries also have hate speech laws targeting the defamation of vulnerable groups, when their members are portrayed as displaying negative traits such as 'depravity, criminality, unchastity, or lack of virtue...[and are therefore exposed]...to contempt, derision, or obloquy' (*Beauharnais v. Illinois* 1952). Other hate speech laws proscribe speech that undermines the victims' dignity, such as 'insults, slurs, evocations of the Ku Klux Klan, statements comparing black men to apes, death threats, and the placement of a noose dangling from the plaintiff's automobile' (*Turley v. ISG Lackawanna Inc.* 2013). Another widely diffuse category of hate speech laws includes laws against incitement to hatred, such as the 2006 Racial and Religious Hatred Act in the UK, which forbids the intentional use of threatening words or behaviour to stir up hatred.

In summary, hate speech regulations may operate at different levels – via formal criminal and civil law at the state level, within more specific institutional contexts (e.g. workplaces, broadcasting companies, etc.), or as part of international conventions and declarations such as the UN International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) (Brown 2015, pp. 39-40; Strossen 2018, p. 108).

## Regulating Hate Speech

According to the United Nations Strategy and Plan of Action on Hate Speech (2019),<sup>4</sup> '[a]ddressing hate speech does not mean limiting or prohibiting freedom of speech. It means keeping hate speech from escalating into something more dangerous, particularly incitement to discrimination, hostility and violence, which is prohibited under international law'. Yet, hate speech laws do inevitably involve restrictions on free speech, and this has generated a longstanding debate among supporters and critics of such laws.

Those who endorse hate speech laws highlight that hate speech harms its victims. For example, it may contribute to short-term harm (e.g. physical harm resulting from incitement to hatred) or to a climate of hatred that may expose its victims to future violence, injustice and discrimination (Brown 2008).

1 [https://www.surveymonkey.com/r/AsianAustralian?fbclid=IwAR1NOQMeSPDMGw3mrBe\\_4DHKOxvq\\_Tz2rJ\\_1Zfx1oPk6M\\_8z2bbWFXnO05lQ](https://www.surveymonkey.com/r/AsianAustralian?fbclid=IwAR1NOQMeSPDMGw3mrBe_4DHKOxvq_Tz2rJ_1Zfx1oPk6M_8z2bbWFXnO05lQ)  
 2 <https://antimuslimhatredworkinggrouphome.files.wordpress.com/2020/04/research-briefing-report-7-1.pdf>  
 3 <https://www.timesofisrael.com/covid-19-fueling-worldwide-wave-of-anti-semitism-researchers-find/>  
 4 <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf>

Victims of hate speech may also suffer from various health conditions, such as anxiety, depression, and high blood pressure (Brown 2015, pp. 56-7). These arguments, however, rely on causal claims that are contestable (Heinze 2016; Strossen 2018). Others contend that hate speech 'inflict[s] psychological injury by assaulting a person's self-respect' (Lawrence 1987, p. 351). According to Waldron (2012), instead, hate speech should be regulated because it undermines its victims' civic dignity, i.e. the 'assurance' that they enjoy an equal social standing as citizens with commonly acknowledged rights and liberties.

Critics of hate speech laws claim that all types of speech, including hate speech, should be allowed in order to allow the truth to emerge from the free circulation of ideas in society (e.g. Chafee 1941; cf. Mill 2006). Yet, when hate speakers, for example, exchange views in echo chambers and internet enclaves, this may reinforce their ill-founded views rather than contributing to truth discovery, and the free circulation of false ideas often results in more people believing in them rather than the opposite (Schauer 2012). More generally, hate speech often hinders, rather than encouraging, the kind of rational deliberation that should help us discover the truth (Brink 2008). It is also unclear why we should prioritize truth discovery over the interests that the victims of hate speech have in their dignity, reputation, and well-being (cf. Schauer 2012).

Other critics of hate speech laws appeal to the value of individual autonomy (cf. Brown 2015, pp. 58-66). C. Edwin Baker (1989, 2009, 2011), for example, argues that autonomous individuals should have the right to make choices and communicate them to others based on their own views and values, including hateful ones. However, hate speech may sometimes undermine, rather than respect or promote, our autonomous agency (Brown 2015, p. 60).

Finally, some critics of hate speech laws argue that such laws undermine democracy in some way, e.g. by limiting citizens' ability to influence public opinion and their political representatives (Post 2011; Weinstein 2011; Heinze 2016). Yet many instances of hate speech do not make any clear contribution to democratic debate; these include, for example, harassment in the workplace, trolling on social media, occasional slurs in the street, and more generally 'the daily low-grade invisible stuff, the hassling, cruel remarks and other things' (Delgado 1991, p. 380 n. 319). Furthermore, the silencing effect of hate speech may often prevent its victims from participating in democratic debate (Brown 2015, p. 198). Ultimately, therefore, both having and not having hate speech laws may undermine democracy in some way (Reid 2020).

### Triggers and Causes of Hate Speech

To understand more about hate speech in times of crisis, we must first understand the motivations that often trigger hate. Primarily, an individual's association as either an ingroup or outgroup member substantially shapes their perceptions towards members of other groups. Intergroup bias, created by ideas of supremacy, identity, or threat, has the potential to erupt into hate and violence (Roussos and Dovidio 2020). Indeed, ingroup members experience less empathy, less trust, and are more fearful towards outgroup members, and may be willing to exploit the weaknesses of others as they perceive members of outgroups to be 'less human' than members of their own group (Leyens et al. 2007 p. 140). Thus, social categorisation plays an important role in intergroup aggression. This 'dehumanisation' is a key component of social prejudice, leading to an 'us' and 'them' mentality that becomes a key component of hate speech and hate crime. Dehumanisation provides a means to reduce empathy towards the other person or group for self-defensive or other instrumental processes (Murrow and Murrow 2015).

It is argued that hate speech is facilitated by seven individual factors that combine to create different degrees of hate

(Sternberg 2020). These include:

1. *fear*, which creates an aversion to, or hatred of, a targeted individual or group;
2. *license*, where a leader grants one the license to hate, overriding any ethical or moral reservations;
3. *obedience to authority*, with more powerful members of society arguing that minorities should be denounced;
4. *trust*, as aggressors increasingly trust their own simplistic and flawed thinking, blame their problems on others, and project their own deficient character onto others;
5. *sense of belonging*, with the hater feeling intimacy and positive commitment toward members of his or her own group but *negation of intimacy* hate toward members of the targeted outgroup;
6. *amplification* of arousal, with manipulative leaders often increasing people's level of arousal against individuals or groups; and
7. *modelling* and imitation of the behaviour of other people, based on observations.

### Hate Speech in Times of Crisis

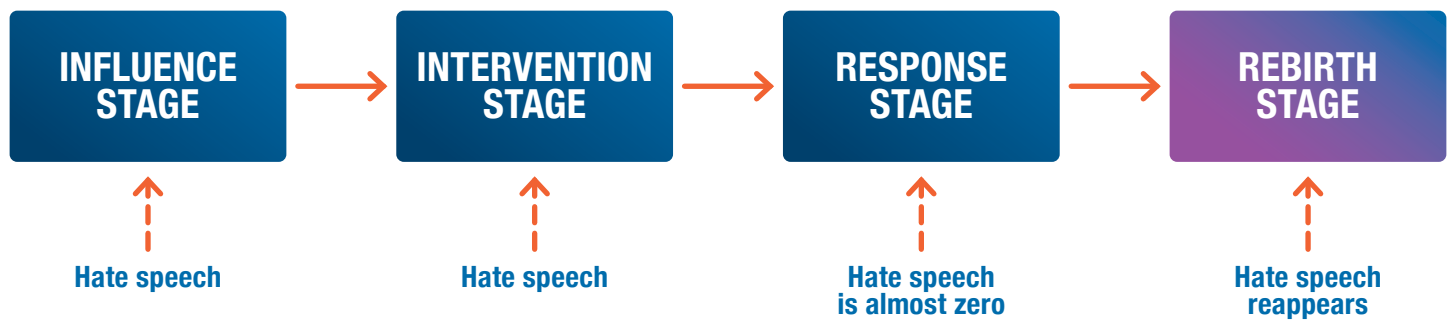
In the empirical literature, there is evidence that discrimination, violence, and exclusion against minorities often increase in times of crisis, such as financial crises, pandemics, emergencies, disasters, wars, and terrorist attacks. While there is some literature on hate speech in times of crisis broadly speaking, crises are so different and varied, each with their own antecedents, that it is difficult to assess the impact of crises on hate speech as a whole. However, there are some clear patterns that emerge.

Incidents that create a significant shock and trauma to society often encourage the dissemination of hate speech (Hinton 2010). The upheaval exacerbates pre-existing anxieties, and focuses people's minds on differences that may be deemed to have contributed to a crisis, such as racial or economic divides 'creating' events such as the 9/11 terrorist attacks or the global financial crisis. The narratives that people employ in response to a cultural upheaval or cultural shock often further divide the world into 'us and them' mentalities. This narrative is also often encouraged by the media through their choice of headlines and language. For example, after the terrorist attacks of 9/11, the media spoke of crusades, a war against terror, and evil doers, evoking images of the medieval campaign Christians waged against the Middle East (Lakoff 2001).

Hate speech is particularly likely to follow a trigger event of terrorism. Chetty and Alathur (2018), for example, argue that hate speech is a staged process: hate speech, immediately after the event (influence stage), will flow heavily on social networks; after a few days (intervention stage) it will reduce; after some more days (response stage) it will reduce to almost zero level; and, after a long time, it may appear once again. This indicates that after a particular event people will be more excited and gradually will return to a normal state or behaviour (Chetty and Alathur 2018, p. 109; see Figure on page 3).

While much research tracks the relationship between hate speech and terrorism, it is important to note that many other types of crisis are important too. For example, times of economic hardship tend to lead to competition between groups rising (Benesch 2014). Similar patterns are also noted in times of political change, and power struggles after the fall of repressive regimes (Benesch 2014). Regardless of the trigger event, we also know that the internet and social media act as amplifiers for hate speech (cf. Brown 2018) and, by allowing people to share or retweet hate-fuelled comments, they may further undermine community relations in times of crisis.

## Stages of hate speech following a trigger event



## REFERENCES

- Baker, C. E. (1989). *Human Liberty and Freedom of Speech* (New York: Oxford University Press).
- Baker, C. E. (2009). 'Autonomy and Hate Speech', in I. Hare and J. Weinstein (eds.), *Extreme Speech and Democracy* (Oxford: Oxford University Press), pp. 139-157.
- Baker, C. E. (2011). 'Autonomy and Free Speech', *Constitutional Commentary* 27 (2): 251-280.
- Benesch, S. (2014). 'Defining and Diminishing Hate Speech', in P. Grant (ed.), *State of the World's Minorities and Indigenous Peoples* (Minority Rights Group International), pp.18-25.
- Brink, D. (2008). 'Mill's Liberal Principles and Freedom of Expression', in C. L. Ten (ed.), *Mill's On Liberty: A Critical Guide* (London: Cambridge University Press), pp. 40-61.
- Brown, A. (2008). 'The Racial and Religious Hatred Act 2006: A Millian Response', *Critical Review of International Social and Political Philosophy* 11 (1): 1-24.
- Brown, A. (2015). *Hate Speech Law: A Philosophical Examination* (New York and Abingdon, Oxon: Routledge).
- Brown, A. (2018). 'What Is So Special about Online (as Compared to Offline) Hate Speech?', *Ethnicities* 18 (3): 297-326.
- Chafee, Z. (1941). *Free Speech in the United States* (Cambridge, MA: Harvard University Press).
- Chetty, N. and S. Alathur (2018). 'Hate Speech Review in the Context of Online Social Networks', *Aggression and Violent Behavior* 40: 108-118.
- Delgado, R. (1991). 'Campus Antiracism Rules: Constitutional Narratives in Collision', *Northwestern University Law Review* 85 (2): 343-387.
- Haraszti, M. (2012). 'Foreword: Hate speech and the Coming Death of the International Standard Before It Was Born', in M. Herz and P. Molnar (Eds.), *The Content and Context of Hate Speech* (New York, NY: Cambridge University Press), pp. xii-xviii.
- Heinze, E. (2016). *Hate Speech and Democratic Citizenship* (Oxford: Oxford University Press).
- Hinton, A.L. 2010. "'Night Fell on a Different World": Dangerous visions and the War on Terror, a Lesson from Cambodia', in A.C.G.M. Robben (ed.), *Iraq at a Distance: What Anthropologists Can Teach Us About the War* (Philadelphia: University of Pennsylvania press) pp 24-57.
- Lakoff, G. (2001). 'Metaphors of Terror'. *UC Berkeley*. Retrieved from <https://escholarship.org/uc/item/8zb6m59x> on 23 July 2020.
- Lawrence, C. (1987). 'The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism', *Stanford Law Review* 39 (2): 317-388.
- Leyens, J. P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M. P. (2007). 'Infra-Humanization: The Wall of Group Differences', *Social Issues and Policy Review* 1 (1): 139-172.
- Mill, J. S. [1859/1869] (2006). *On Liberty and The Subjection of Women* (ed. A. Ryan) (London: Penguin Books).
- Murrow, G. B. and Murrow, R. (2015). 'A Hypothetical Neurological Association Between Dehumanization and Human Rights Abuses', *Journal of Law and the Biosciences* 2 (2): 336-364.
- Post, R. (2011). 'Participatory Democracy and Free Speech', *Virginia Law Review* 97 (3): 477-89.
- Reid, A. (2020). 'Does Regulating Hate Speech Undermine Democratic Legitimacy? A Cautious "No"', *Res Publica* 26 (2): 181-199.
- Roussos, G. and Dovidio, J. F. (2020). *Tolerating hate: Racial bias, freedom of speech, and responses to hate crimes*, in R. J. Sternberg (ed.), *Perspectives on Hate: How It Originates, Develops, Manifests, and Spreads* (American Psychological Association), pp. 225-249. Retrieved from <https://doi.org/10.1037/0000180-011> on 23 July 2020.
- Schauer, F. (2012). 'Social Epistemology, Holocaust Denial, and the Post-Millian Calculus', in M. Herz and P. Molnar (eds.), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), pp. 129-144.
- Sternberg, R. J. (2020). 'FLOTSAM: A Theory of the Development and Transmission of Hate', in R. J. Sternberg (ed.), *Perspectives on Hate: How It Originates, Develops, Manifests, and Spreads* (American Psychological Association), pp. 3-24. Retrieved from <https://doi.org/10.1037/0000180-001> on 23 July 2020.
- Strossen, N. (2018). *Hate: Why We Should Resist It with Free Speech, Not Censorship* (Oxford: Oxford University Press).
- Waldron, J. (2012). *The Harm in Hate Speech* (Cambridge, Mass.; London: Harvard University Press).
- Weinstein, J. (2011). 'Participatory Democracy as the Central Value of American Free Speech Doctrine', *Virginia Law Review* 97 (3): 491-514.

## LEGAL CASES CITED

- Beauharnais v. Illinois*, 343 U.S. 250 (1952)
- Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942)
- EEOC v. Tyson Foods, Inc.*, cv-05-BE-1704-E (N.D. Ala. Nov. 7, 2006)
- Turley v. ISG Lackawanna, Inc.*, No. 13-561 (2d Cir. 2014)
- Virginia v. Black*, 538 U.S. 343 (2003)