

1 Approaches to the meaning of hate speech

1 Introduction

The term *hate speech* recognises a wide range of harmful social practices and discourses. Let us take, for instance, Islamophobic blogs, cross burnings, racial epithets, or dehumanising pictures of Jews. If all these social activities fall under hate speech, they must have certain elements or features in common. It might therefore be reasonable to think that a definition of hate speech should comprise the essential elements that enable us to recognise the phenomenon. This chapter aims to demonstrate how Wittgenstein's concept of *family resemblance*, which Brown⁷ borrowed in his legal essay *What is hate speech? Part II: Family resemblances* (2017b), can help linguists and legal scholars better describe hate speech. The discussion is divided into three parts, each corresponding to a different approach to hate speech. The first approach I deal with is *ordinary language analysis*, a philosophical investigation method concerned with how verbal expressions, in our case, the term hate speech, is used in non-technical, everyday language (Wittgenstein 2009 [1953]). Then I analyse various legal scholarly attempts to define hate speech. Finally, the chapter provides an overview of the approaches to a technical legal definition of hate speech in international law, common law and civil law.

2 Wittgenstein's concept of *family resemblance* applied to the definition of hate speech

Family resemblance is a philosophical concept found in Wittgenstein's post-humously published *Philosophical investigations* (2009 [1953]). Wittgenstein turned to games, as a way to explain the gist of his idea:

Look, for example, at board games, with their various affinities. Now pass to card games; here, you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball games, much that is common is retained, but much is lost [...] And the upshot of these considerations is that we see a complicated network of similarities overlapping and crisscrossing; similarities in the large and in the small (Wittgenstein 2009 [1953]: 36).

⁷ Alexander Brown is a Reader in Political and Legal Theory at the University of East Anglia in the United Kingdom.

The application of the concept of family resemblance to the various definitions of hate speech involves, on the one hand, comparing disparate definitions and identifying sets of overlapping and crisscrossing similarities between those definitions. The essential idea is that any definition of hate speech shares at least one similar element with at least one other definition, even if there is no single element common to all definitions. As a result, as Brown argued, “[. . .] we are likely to end up with a family of different meanings of the term hate speech, each slightly different from the next but sharing at least one essential element in common” (Brown 2017b: 13).

3 Ordinary language analysis: An approach to understanding the meaning of hate speech in everyday language

Wittgenstein (2009 [1953]) laid the groundwork for ordinary language analysis: he advocated that, in solving philosophical problems, understanding how language is used in context is more important than its abstract meaning. In “What is hate speech? Part 2: Family resemblances”, Brown (2017b) applies ordinary language analysis to describe how the term hate speech is used in non-technical, everyday language. He divides his analysis into four parts: (1) *purpose-oriented analysis*, (2) *folk platitudes analysis*, (3) *intuitions analysis* and (4) *everyday use analysis*. In the following, I summarise the major insights Brown provides in each type of analysis.

- 1) Purpose-oriented analysis. The term hate speech may be used for five different purposes in non-technical, everyday language: (a) highlighting forms of harmful speech, (b) pointing out socially disruptive forms of speech, (c) identifying forms of speech that can challenge people’s sense of equality, (d) articulating norms of civility and (e) labelling forms of speech that undermine democracy.
- 2) Folk platitudes analysis. There seem to be four main commonplace generalisations, which Brown (2017b) calls folk platitudes, held by people about the meaning of the term hate speech: (a) it is negatively evaluative; (b) it is directed against groups of people identifiable by legally-protected characteristics – e.g. race, ethnicity, religion, sexual orientation, gender identity and disability; (c) it relates to emotions, feelings, or attitudes that can lead to incitement to hatred, and are even liable to trigger violence; and d) it is speech that is not protected by an individual’s right to freedom of opinion and expression.
- 3) Intuitions analysis. People seem to have certain intuitions about the forms of speech that may be associated with hate speech. Most people would agree

on classifying the following forms of speech as hate speech when they are addressed to historically oppressed or victimised groups: (a) insults, racial epithets and ethnic jokes (Leader, Muller & Rice 2009); (b) any forms of speech that articulate ideas relating to the moral inferiority or non-humanity; (c) presenting false statements that harm social reputation; and (d) words or behaviour that threaten, support, or incite hatred or violence, or words that simply justify or glorify violence.

- 4) Everyday language. There seem to be several speech acts and discourses that give expression to hate speech, such as racial hate speech and homophobic hate speech.

In sum, Brown's (2017b) ordinary language analysis approach to the meaning of the term hate speech demonstrates that in non-technical, everyday language, the term does not have a single meaning but, instead, a family of meanings, each slightly different from another but sharing at least one essential element in common.

4 Legal scholarly attempts to defining the concept of hate speech

Some significant legal scholarly attempts to define hate speech began in the 1980s and continue today. Marwick and Ross argued that one can broadly define hate speech as “[. . .] speech that carries no meaning other than hatred towards a particular minority, typically a historically disadvantaged minority” (Marwick & Ross 2014: 17). Marwick and Ross identified three distinct categories that legal scholars use to define hate speech:

- 1) Content-based hate speech includes words, expressions, symbols and iconographies, such as the Nazi swastika or the Ku Klux Klan's burning cross, generally considered offensive to a particular group of people and objectively offensive to society.
- 2) Intent-based hate speech requires the speaker's communicative intention to incite hatred or violence against a particular minority, member of a minority, or person associated with a minority without communicating any legitimate message.
- 3) Harms-based hate speech is speech that causes the victim harm, such as loss of self-esteem, physical and mental stress, social and economic subordination and effective exclusion from mainstream society.

In the following sections, I will review some of the most significant definitions in the legal scholarly domain and classify them according to the above categories or elements identified by Marwick and Ross (2014).

4.1 Content-based hate speech

In the late 1980s, a group of legal scholars in the United States used the term “racist speech” in response to how different legal systems tackled a particular type of harmful speech targeted against specific racial groups. In his influential work, *Words that wound: A tort action for racial insults, epithets, and name-calling* (1982), Delgado opened the path to scholarly discussion about why racist speech should be considered a civil tort under US law. Delgado (1982: 179) proposed an element-based definition that requires racist speech to be actionable if: (1) the language that was addressed to the plaintiff by the defendant intended to demean through reference to race, (2) the plaintiff understood the language as intended to demean through reference to race and (3) a reasonable person would recognise the language employed by the defendant as a racial insult. Delgado, then, defined racist speech as a specific type of civil tort. Accordingly, his definition comprised the three essential elements in a civil tort, with the only difference being that the offence revolves around insult on racial grounds: (1) an intent-based element to cause harm through racial insult, (2) the plaintiff must secure uptake of the intended racial insult and (3) the act must be, objectively, a racial insult.

In the same vein, Matsuda (1989) claimed that US law should provide legal means for bringing action against hate-advocating speakers in America by criminalising racist speech because of its harmful effects on the targeted racial minorities and on democratic social values. Matsuda’s standpoint had as its base the provisions of international law such as the Universal Declaration of Human Rights,⁸ the International Covenant on Civil and Political Rights (ICCPR)⁹ and the International Convention on the Elimination of All Forms of Racial

⁸ The United Nations General Assembly proclaimed the Universal Declaration of Human Rights as a common standard of achievements for all peoples and all nations in Paris on 10 December 1948 (General Assembly resolution 217 A). <https://www.un.org/en/universal-declaration-human-rights/> (accessed 31 May 2020).

⁹ The International Covenant on Civil and Political Rights. Adopted by the General Assembly of the United Nations on 19 December 1966. <https://treaties.un.org/doc/publication/unts/volume%20999/volume-999-i-14668-english.pdf> (accessed 26 May 2020).

Discrimination (CERD).¹⁰ It should be noted that Matsuda did not make any distinctions between racist speech, hate speech, hateful speech and hateful messages. She used the terms interchangeably to refer to the same social phenomenon. Although, in Matsuda's work, the term hate speech primarily refers to racism, she also recognised other types of hate speech based on sex and sexual orientation, such as anti-gay and anti-lesbian hate speech, that, in her judgement, require independent attention. For Matsuda, racial hate speech and hate speech on the grounds of sexual orientation are products of different types of discourses. Specifically, whereas racial hate speech is a product of a discourse of subordination that claims the superiority of the white race over the black race, hate speech on the grounds of sex and sexual orientation is a product of a discourse of power that claims the supremacy of masculinity and heterosexuality over femininity and homosexuality. Matsuda (1989: 2358) suggested that there are three elements at the core of hate speech: (1) the message is of racial inferiority, (2) the message is directed against a historically oppressed racial group and (3) the message is hateful and degrading. Unlike the definition of Delgado, Matsuda's seems to be more effect-oriented in the first and third elements of the definition – that is, “the message is of racial inferiority”, “the message is hateful and degrading”. Matsuda's definition also highlights that hate speech must be “directed against historically oppressed groups”.

In her work, Matsuda considered the violence of the word in *threats, racial epithets*¹¹ and *slurs*,¹² and the violence of *symbols*, such as the Klan burning crosses and their robes, which take their hateful meaning from their historical context and connection to violence. Matsuda's approach foregrounded the damaging effects hate speech may have on social peace because it is a speech that causes hatred, persecution and degradation of certain groups.

Parekh (2006) drew attention to the difficulty of recognising hate speech because of its close and complex association with conflict¹³ (Nelde 1987:

10 The International Convention on the Elimination of All Forms of Racial Discrimination. Adopted and opened for signature and ratification by General Assembly resolution 2106 (XX) of 21 December 1965, entry into force 4 January 1969, under Article 19. <https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx> (accessed 3 June 2020).

11 A racial epithet, also known as a racial slur, is a derogatory term or expression based on someone's racial background. Racial epithets are understood to convey contempt and hatred toward their targets (cf. Hom 2008).

12 A slur is an insulting remark that could damage someone's reputation.

13 Language conflict involves intrastate political tension or civil unrest between speakers of different lingua-cultures. Language conflict plays a central role in group and national identity; hence, it is a surface indicator of entrenched political, social and economic conflict.

33–42; Darquennes 2015: 7–32; Millar 2019: 145–163), offence, abuse and discrimination (Stollznow 2017). Hate speech can be triggered by social conflict, but hate speech itself can also spark conflict, violence and a breach of the peace. Hate speech can be conveyed through offensive,¹⁴ abusive¹⁵ and other types of harmful speech. It must be clarified that these types of speech, which should be considered broad categories including many types of rude, nasty, malicious or even humiliating expressions, are only manifestations of hate speech under specific circumstances – e.g. the hateful message must be addressed to a legally-protected group.

Parekh argued that the reasons for prohibiting hate speech are specific and should be distinguished from those restricting other types of language crimes, such as defamation (libel and slander) and threats. Parekh (2006: 214) proposed a definition of hate speech based on three essential elements: (1) it singles out an individual or a group based on legally-protected characteristics, (2) it stigmatises the target group and (3) it marginalises the target group. Parekh's definition focused on the discriminatory content hate messages convey, as it stigmatises, de-humanises and demonises the members of the target group with the intent of placing them outside of public life. For Parekh, it is the hateful content of the messages, rather than the immediate consequences, that should matter to the law. In Parekh's words:

It is, therefore, a mistake commonly made to define hate speech as one likely to lead to public disorder and to proscribe it because or only when it is likely to do so. What matters is its content, what it says about an individual or a group, not its likely immediate consequences, and our reasons for banning it need not be tied to the latter (Parekh 2006: 214).

Since Parekh's definition does not provide an intent-based element, one can expect it to attract severe criticism from freedom of speech theorists. Apart from pointing to the harmful content of hate speech, Parekh (2006: 214) also laid stress on the multimodal nature and pragmatic meaning of hate speech forms. Parekh argued that although speakers often convey hatred through explicit, offensive and abusive language, hate speech can also be expressed in implicit and moderate language. For instance, hate speech can be embedded in maliciously ambiguous jokes, implicatures and non-verbal elements of communication (see Chapter 8). From Parekh's insights, one can learn that although a violent mode of expression provides a useful clue to the malevolent nature of a

14 Offensive language refers to hurtful, derogatory or obscene comments by one person to another.

15 Abusive language refers to harsh, violent, profane, or derogatory language that could demean the dignity of an individual.

message, it is inadequate to take overtly violent messaging as the only clue. One should also be cautious in associating hate speech with abusive language, because this type of negative language can also be used to describe other language crimes, such as insults, defamation (libel or slander) and threats. For Parekh (2006), the meaning of hate speech can only be understood in the context of situation. An utterance may appear to be innocent in one context, while the same utterance may not be so innocent in another context. This ambiguous condition would be a hindrance to the elaboration of a closed catalogue or inventory of hate speech forms.

Waldron (2012) also proposed a content-based definition of hate speech. In his view, hate speech refers to group defamation or group libel because the messages may involve several reputational attacks that amount to “assaults upon the dignity of the person affected” (Waldron 2012: 59). For Waldron, a message may qualify as hate speech if it involves “some imputation of terrible criminality” (Waldron 2012: 47) that can affect the human dignity of people who belong to a legally-protected group.

4.2 Intent-based hate speech

Moran (1994) defined hate speech as “speech that is intended to promote hatred against traditionally disadvantaged groups” (Moran 1994: 1430). Moran’s definition introduced a new element: the perpetrator’s malicious intent to “promote hatred”, and it also broadened the scope of possible target groups because the wrong action is not exclusively directed against racial groups. Ward (1997), for his part, claimed that the legal restraint of hate speech is a means to strengthen the individual’s right to freedom of opinion and expression. He defined hate speech as “any form of expression through which speakers intend to vilify, humiliate, or incite hatred against their targets” (Ward 1997: 765). His definition includes three elements: (1) the subject’s intent to damage the target’s rights to human dignity and equality, (2) the subject’s intent to incite hatred, which involves urging or persuading someone to act violently or unlawfully against the target and (3) the message must be perceived as a *verbal attack* against the target groups.

In the same line of thought, Benesch described a specific subcategory of hate speech that she named *dangerous speech* (Benesch 2014: 22) because of its power to call for violence against the target groups. Benesch’s dangerous speech resembles Matsuda’s *fighting words* (1989: 2355) and Ward’s *verbal attack* (1997: 766) in that the term foregrounds the element of intent to incite hatred or violence against the targets.

4.3 Harms-based hate speech

Drawing on Matsuda's (1989) definition of racist speech, Massey (1992) provided a broad and detailed argument on the legal reasons for suppressing hate speech based on the foundational paradigms of free speech. Massey's definition of hate speech exclusively focuses on the harm it causes to freedom of opinion and expression. In his view, hate speech should be actionable if it (1) causes harm or violence to individuals belonging to specific groups and (2) abuses the right to freedom of opinion and expression by inflicting injury on and causing despair to its targets. In essence, Massey's definition shares with Matsuda's one main idea: the principles of racist ideology are irreconcilable with the foundational paradigms of equality and freedom of speech of the Constitution of the United States. For this reason, according to Massey, hate speech should be considered restricted speech and regulated by law.

After reviewing the three main legal scholarly approaches to the definition of hate speech, content-based, intent-based and harms-based, it is clear that hate speech is not a univocal concept that can be defined unambiguously, but rather a multi-layered concept with multiple meanings that requires various approaches. As a result, one has a family of legal scholarly definitions of hate speech, each slightly different from the other in perspective but sharing at least one core element with another definition. Therefore, I can also conclude that Wittgenstein's concept of family resemblance (2009 [1953]) improves our understanding of the various legal scholarly approaches to the definition of hate speech.

5 Approaches to a technical legal definition of hate speech

Since World War II, many nations have passed laws to curb incitement to hatred against minorities. These laws, which started as protections against anti-Semitism, are known under the rubric of hate speech laws. The elements of hate speech, as in the case of specific civil or criminal offences, must be captured in their legal definition, which enables prosecutors, judges and juries to examine the offence in light of the specific circumstances of a given legal case. However, the problem, as the next sections show, is that there is no agreed-upon technical legal definition of hate speech. The discussion that follows then aims to provide good examples of the differing approaches to a technical legal definition of hate speech at three levels: international law, common law and civil law – European Union law and Member State law.

5.1 International law

The Universal Declaration of Human Rights (UDHR) (1948) is at the core of hate speech laws because it set out the principle, for the first time, that there must be universal protection for people's fundamental rights. Since hate speech involves vicious acts against human dignity and equality, it violates the fundamental rights articulated in Article 2:

Everyone is entitled to all the rights and freedoms outlined in this Declaration, without distinction, such as race, colour, sex, language, religion, political opinion, national or social origin, property, birth or status. Furthermore, no distinction shall be made based on the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, trust, non-self-governing or under any other limitation of sovereignty.

In the same vein, the International Convention on the Elimination of All Forms of Racial Discrimination (CERD) (1965) imposes responsibility on states for the prohibition of hate speech by law in Article 4:

States Parties condemn all propaganda and organisations disseminating ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights.

Hate speech must be prohibited by law because it infringes on universal fundamental rights, such as the rights to human dignity and equality. The International Covenant on Civil and Political Rights (ICCPR) (1966) refers to hate speech in Article 20, paragraph 2), as

Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited.

Hate speech, then, consists of three essential elements that must co-occur:

- 1) Advocacy: Public forms of expression intended to elicit action or response.
- 2) Hatred: Intense negative emotions towards a group of people identified by legally-protected characteristics.
- 3) Incitement: Public forms of expression likely to trigger acts of discrimination, hostility or violence against a group of people identified by legally-protected characteristics.

Although there is no agreed-upon technical legal definition, both the International Convention on the Elimination of All Forms of Racial Discrimination

(CERD) (1965) and the International Covenant on Civil and Political Rights (ICCPR) (1966) share three elements in their definition of hate speech: (1) perpetrators use public forms of expression, (2) perpetrators target victims belonging to a historically disadvantaged group of people and (3) perpetrators have a malicious intent. Both the International Convention on the Elimination of All Forms of Racial Discrimination (CERD) (1965) and the International Covenant on Civil and Political Rights (ICCPR) (1966) identify the element of intent. However, their approaches to hate speech are different. Whereas the former refers to a wider array of speech acts, such as “disseminate”, “justify”, “promote” and “incite” racial hatred and discrimination against the targets, the latter unambiguously focuses on the intent “to trigger acts of discrimination, hostility or violence” and its scope is not limited to racial hatred.

5.2 Common law

5.2.1 Hate crime legislation in the United States

At the federal level, the Hate Crime Statistic Act does not refer to hate speech but instead to hate crimes in the following terms:

Under the authority of section 534 of title 28, United States Code, the Attorney General shall acquire data, for each calendar year, about crimes that manifest evidence of prejudice based on race, gender and gender identity, religion, disability, sexual orientation, or ethnicity, including where appropriate the crimes of murder, nonnegligent manslaughter; forcible rape; aggravated assault, simple assault, intimidation; arson; and destruction, damage or vandalism of property.¹⁶

US law therefore prohibits hate crimes against individuals who belong to legally-protected groups. Unlike hate crime, hate speech does not have a legal definition under US law and enjoys substantial protection under the First Amendment to the Constitution of the United States. According to the First Amendment:

Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press, or the right of the people peaceably to assemble and to petition the government for a redress of grievances.¹⁷

¹⁶ Hate Crime Statistics Act, As Amended, 28 USC. § 534 § [Sec. 1.]. <https://ucr.fbi.gov/hate-crime/2017/resource-pages/hate-crime-statistics-act.pdf> (accessed 27 May 2020).

¹⁷ The Constitution of the United States. First Amendment. <https://constitution.congress.gov/constitution/amendment-1/> (accessed 27 May 2020).

The First Amendment to the Constitution of the United States protects speech no matter how offensive its content may be. This protection entails from the assumption that the fundamental right to freedom of opinion and expression requires the government to strictly protect vigorous debate on matters of public concern even when such debate devolves into a dangerous speech that causes other people to feel grief, anger, or fear. This does not mean that the First Amendment does not provide any restrictions to hate speech. It does, in effect, impose limitations on forms of expression such as “genuine threats”, “defamatory speech”, “obscenity” and “incitement to imminent violence”. An individual’s speech may be restricted in the United States if (1) it is intended to produce lawless action, (2) it is likely to incite such action and (3) such action is likely to occur imminently. This set of restrictions to freedom of opinion and expression stem from the case of *Brandenburg v. Ohio* (1969), whose final judgment is known for having replaced the *clear and present danger standard* with the *imminence standard*. In this case, *Brandenburg*, a Ku Klux Klan leader, was convicted by the Court in Ohio for delivering a protest speech at a Klan rally, in which he advocated violence against African Americans and Jews. The Supreme Court did not only overturn *Brandenburg*’s conviction but also struck down the Ohio statute prohibiting his speech because

Freedoms of speech and press do not permit a State to forbid advocacy of the use of force or law violation except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action (*Brandenburg v. Ohio* (1969). No. 492).¹⁸

As shown in the above excerpt from the Supreme Court’s judgment in *Brandenburg v. Ohio* (1969), hate speech in US law strictly refers to that speech that “is directed to inciting or producing imminent lawless action, and it is likely to incite or produce such action.” In practice, this definition imposes a very high standard for a case of hate speech to be declared in US courts of justice. The definition also confirms the protected status of the right to freedom of opinion and expression under the First Amendment to the Constitution of the United States.

5.2.2 Hate speech legislation in Canada

Under Canadian law, hate speech refers to speech that publicly and wilfully promotes hatred against groups of people that can be identified by legally-

18 US Supreme Court. *Brandenburg v. Ohio*, 395 US 444 (1969). *Brandenburg v. Ohio*. No. 492. Argued February 27, 1969. Decided June 9, 1969. <https://supreme.justia.com/cases/federal/us/395/444/> (accessed 18 August 2020).

protected characteristics, as stated in section 318(4) of the Criminal Code.¹⁹ In addition, at a provincial level, most provinces contain provisions in their respective human rights legislation that prohibit forms of expression that incite hatred or violence against individuals or groups identified by legally-protected characteristics.

5.2.3 Hate speech legislation in the United Kingdom

Although there is no specific hate speech law in the United Kingdom, several laws forbid incitement to hatred or violence against legally-protected groups. The hate crime legislation in the United Kingdom criminalises conduct that is intended or likely to stir up racial hatred or is intended to do so based on religion or sexual orientation. The term “conduct” includes the use of hateful words and a wide range of forms of expression, such as displays of text, books, banners, photos and visual art, the public performance of plays and the distribution or presentation of pre-recorded material. The legal concept of “stirring up offences” is an effort to challenge racial hatred in the United Kingdom. Racial hatred, which was initially enacted in the Race Relations Act 1968,²⁰ is defined in the Public Order Act 1986 as “hatred against a group of persons defined by reference to colour, race, nationality (including citizenship) or ethnic or national origins.”²¹ Sections 18 to 22 of the Public Order Act 1986 set out several acts through which racial hatred is likely to be stirred up. For example, by using words, behaviour or displaying written material (s. 18), publishing or distributing material (s. 19), publicly performing a play (s. 20), distributing, showing or playing a recording (s. 21), broadcasting or including a programme in cable programme service (s. 22).

Over the last few years, new incitement offences have joined the Public Order Act 1986 – e.g. religious hatred joined in 2006 and incitement to hatred on sexual orientation grounds joined in 2008. The legislation defines religious hatred as “hatred against a group of persons defined by religious belief or lack of religious belief”,²² and hatred on the grounds of sexual orientation as “hatred against a group of persons defined by sexual orientation (whether towards persons of the

¹⁹ Criminal Code, R.S.C. 1985, c.C-46, s. 318(4). <https://laws-lois.justice.gc.ca/eng/acts/c-46/page-68.html#docCont> (accessed 27 May 2020).

²⁰ Race Relations Act 1968 c.71. Part I. Discrimination. <http://www.legislation.gov.uk/ukpga/1968/71/enacted> (accessed 27 May 2020).

²¹ Public Order Act 1986. Part III. Racial Hatred, s.17. Meaning of racial hatred. <http://www.legislation.gov.uk/ukpga/1986/64/contents> (accessed 27 May 2020).

²² Public Order Act 1986. Part III. Racial Hatred, s. 29A. Meaning of religious hatred. <http://www.legislation.gov.uk/ukpga/1986/64/contents> (accessed 27 May 020).

same sex, the opposite sex or both sex)".²³ Sections 29B to 29F set out acts intended to stir up religious hatred or hatred on the grounds of sexual orientation. These acts are the same as those set out for racial hatred.

5.2.4 Hate speech legislation in Australia

In Australia, the Racial Discrimination Bill 1975 (RDB) declared the prohibition of offensive behaviour motivated by others' race, colour or national or ethnic origin. Specifically, section 18C states:

It is unlawful for a person to do an act, otherwise than in private, if:

- 1) the act is reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate another person or a group of people; and
- 2) the act is done because of race, colour or national or ethnic origin of the other person or some or all people in the group.²⁴

Another relevant law in Australian legislation in which one can find an attempt to identify the punishable elements of hate speech is the Criminal Code Amendment (Racial Vilification Act), which is a hate speech law that uses the terms "racial animosity" and "racist harassment" rather than hate speech. The Racial Vilification Act describes unlawful conduct in sections 77–80D, ranging from conduct intended to incite hatred against targeted racial groups, dissemination of material for incitement to racial hatred and racial harassment. Specifically, section 77 defines conduct intended to incite racial animosity or racial harassment as

Any person who engages in any conduct, otherwise than in private, by which the person intends to create, promote or increase animosity towards, or harassment of, a racial group or a person as a member of a racial group is guilty of a crime and is liable to imprisonment for 14 years.²⁵

The above definition refers to three essential elements of racial hate speech: (1) it involves malicious conduct directed against a racial group, (2) it requires a specific intent to commit the crime and (3) the purpose of this type of harmful conduct is to incite racial animosity or racial harassment towards a racial group.

²³ Public Order Act 1986. Part III. Racial Hatred, s. AB. Meaning of hatred on the grounds of sexual orientation. <http://www.legislation.gov.uk/ukpga/1986/64/contents> (accessed 27 May 2020).

²⁴ The Racial Discrimination Bill 1975. <https://www.legislation.gov.au/Details/C2014C00014> (accessed 28 May 2020).

²⁵ The Racial Vilification Act 2004. Section 77. https://www.legislation.wa.gov.au/legislation/statutes.nsf/law_a9290_currencies.html (accessed 28 May 2020).

One significant difference between common law legal systems is the specific weight they give to the right to freedom of opinion and expression over the right to dignity and equality. On the one hand, the hate speech legislation in the United Kingdom and Australia seem to favour the fundamental rights to dignity and equality over the right to freedom of opinion and expression. UK law prohibits speech that stirs up hatred on the grounds of race, religion and sexual orientation, while Australian law prohibits racial animosity and racial harassment. In these two common law jurisdictions, hate speech is related to the incitement to hatred, especially racial hatred or racial harassment, and it is this type of restricted speech that the law prohibits without considering whether it may, or may not, lead to violent acts against the target group. On the other hand, hate speech laws in the United States and Canada seem to favour the right to freedom of opinion and expression over the right to dignity and equality when they only ban speech that is “likely to incite or produce imminent lawless action” (US law) or is “likely to cause a breach of the peace” (Canadian law).

5.3 Civil law

5.3.1 Hate speech in European Union law

Hate-speech legislation in Europe goes back to the European Convention on Human Rights.²⁶ Specifically, Article 14 prohibits discrimination based on sex, race, colour, language, religion, political opinion, national or social origin, association with a national minority, property, birth or status. The fundamental rights to dignity and equality prevail over the right to freedom of opinion and expression under European Union law. Recommendation No. R (97) 20 of the Committee of Ministers of the Council of Europe to the Member States, adopted on 30 October 1997, defines hate speech as

all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin (Committee of Ministers of the Council of Europe 1997: 107).²⁷

²⁶ European Convention on Human Rights. Convention for the Protection on Human Rights and Fundamental Freedoms. Rome, 4. XI, 1950. https://www.echr.coe.int/Documents/Convention_ENG.pdf (accessed 27 May 2020).

²⁷ Council of Europe Recommendation No. R (97) 20 of the Committee of Ministers to the Member States on Hate Speech (Adopted by the Committee of Ministers on 30 October 1997, at

The above definition presents two important differences compared with the common law. First, its scope is broader because, apart from “incitement to hatred”, it considers other types of acts, such as “spread”, “promote” and “justify”. Second, the target groups are more varied, as they refer to racial hatred and xenophobia, anti-Semitism and other hate-based intolerance against minorities and migrants. In the same vein, the Council Framework Decision (2008),²⁸ on combating certain forms and expressions of racism and xenophobia through criminal law, defines hate speech as certain forms of conduct which are punishable as criminal offences, such as (1) public incitement to violence or hatred directed against a group of persons or a member of such a group on the grounds of race, colour, descent, religion or belief, or national or ethnic origin; (2) public dissemination or distribution of pictures or other material; (3) publicly condoning, denying or grossly trivialising crimes of genocide, crimes against humanity and war crimes; and (4) instigating, aiding or abetting in the commission of the offences mentioned in (1), (2) and (3).

The definition provided by the Council Framework Decision (2008) comprises a wider range of forms of expression, apart from “incitement to violence or hatred”, such as “dissemination or distribution of materials”, “condoning”, “denying” or “trivialising crimes of genocide, crimes against humanity and war crimes”, “instigating”, “aiding” and even “abetting” in the commission of hate speech offences. Nevertheless, it does not broaden the scope of the target groups, as it focuses on racial and religious hatred.

Following the recommendations of the European Commission against Racism and Intolerance (ECRI), the Council of Europe defines hate speech as

the advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatisation or threat in respect of such a person or group of persons and the justification of all the primary types of expression, on the ground of race, colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status.²⁹

the 607th Meeting of the Ministers’ Deputies. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680505d5b> (accessed 27 May 2020).

28 Council Framework Decision 2008/913/JHA of 28 November 2008. Entry into force 6 December 2008. *Official Journal* OJ L 328 of 6 December 2008. This decision ensures that specific severe manifestations of racism and xenophobia are punishable by effective, proportionate and dissuasive criminal penalties throughout the European Union.

29 Council of Europe. No Hate Speech Movement. <https://www.coe.int/en/web/venice/hate-speech-movement> (accessed 27 May 2020).

In the definition above, one can find two core elements common to definitions of hate speech: public and wilful incitement to hatred directed against a legally-protected group of people. However, the definition is ambiguous for various reasons. Firstly, the use of the disjunctive conjunction “or” makes legal interpretation difficult because the options are presented as mutually exclusive when, in effect, they may not be exclusive. Secondly, the definition refers to a wide range of acts – e.g. “advocacy”, “promotion”, “incitement”, “justification”, “harassment”, “insult”, “threat”, and perlocutionary effects – e.g. “denigration”, “vilification”, “negative stereotyping” and “stigmatisation”. Thirdly, the content points to a broad spectrum of legally-protected characteristics: “race”, “colour”, “descent”, “national or ethnic origin”, “age”, “disability”, “language”, “religion”, “belief”, “sex”, “gender”, “gender identity” and “sexual orientation”. Fourthly, the definition contains vague references, such as “other personal characteristics or status”, that, in practice, would be difficult for legal practitioners to elucidate.

5.3.2 Member State law

This section illustrates different approaches to a technical legal definition of hate speech in Member State law. Due to spatial constraints, the analysis focuses on three EU Member States representing various legal cultures within European civil law: German civil law, French civil law and Spanish civil law.

a) Hate speech in German legislation

Although German legislation does not have a specific hate speech law, there are several penal, civil and administrative provisions against the most severe hate speech and hate crimes. The primary piece of criminal legislation prohibiting hate speech in Germany is the Criminal Code. This law provides a distinction between bias-motivated crimes (*Vorurteilsdelikte*) and symbolic crimes (*Botschaftsverbrechen*). With respect to bias-motivated crimes, section 46, paragraph 2), of the Criminal Code states that

racist, xenophobic and other inhumane or contemptuous motives are aggravating circumstances to be considered when establishing the grounds for sentencing for any crime under the Criminal Code.³⁰

³⁰ Criminal Code in the version published on 13 November 1998 (Federal Law Gazette I, p. 3322), as last amended by Article 2 of the Act of 19 June 2019 (Federal Law Gazette I, p. 844). Section 46. The German Federal Ministry of Justice provides a translation into English of the official German texts on their website. https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html#p1333 (accessed 29 May 2020).

In this respect, several provisions in the German Criminal Code are applicable in legal cases involving instances of hate speech, such as insult (under section 185), malicious gossip (under section 186), defamation (under section 187), malicious gossip and defamation concerning persons in political life (under section 188) and defiling the memory of the dead (under section 189). These provisions are applicable in both group defamation and defamation against a member of the group, in which there is bias motivation – that is, when the offences in question target a group or an individual that belongs to a group that can be identified by legally-protected characteristics.

On the other hand, symbolic crimes in Germany explicitly include “incitement to hatred”. Specifically, in section 130, Incitement of masses, of the Criminal Code, one can find two elements of hate speech: (1) speech that relates to incitement to hatred and (2) speech that is directed against people who belong to a legally-protected group. The German Criminal Code also describes the forms of expression of hate speech. These include: “incitement to hatred and violence”, “violation of the human dignity of others by disseminating, insulting, dehumanising and defaming”. In sum, German law seems to favour the right to dignity over the right to freedom of opinion and expression when the exercise of the latter is considered abusive.

b) Hate speech in French legislation

Hate speech in France is prohibited under both criminal law and civil law. The French Penal Code forbids any private defamation of a person or group of people for belonging to a group that can be identified by legally-protected characteristics (Article 132–176):

The aggravating circumstances defined in the first paragraph are applicable when the offence is preceded, accompanied or followed by written or spoken words, images, objects or actions of whatever nature that damage the honour or the reputation of the victim, or a group of persons to which the victim belongs, on account of their actual or supposed membership or non-membership of a given ethnic group, nation, race or religion.³¹

31 Penal Code. English translation. With the participation of John Rason Spencer QC, Professor of Law, University of Cambridge, Fellow of Selwyn College. Article 132–76. Updated 12 October 2005, 49/132. file:///C:/Users/Victoria%20Guillen/Downloads/Code_34%20(2).pdf (accessed 29 May 2020).

Apart from the criminal provisions in the Penal Code, Article 23 of the Law on the Freedom of the Press of 29 July 1881 (as of 2014)³² prohibits discrimination based on origin, membership, or non-membership of a racial or religious group. Moreover, Articles 32– 33 in the same law prohibit anyone from defaming or insulting a person or group identified by legally-protected characteristics. In addition, Article 34 prohibits defiling the memory of the dead. Therefore, in French legislation, hate speech is speech that defames, insults, or defames the memory of members of legally-protected groups of people based on ethnicity, nationality, race and religion. More recently, the categories of sex, sexual orientation and disability have been added to the already existing legally-protected characteristics.

c) Hate speech in Spanish legislation

As in the case of Germany and France, Spain has no specific hate speech law, nor does it have a specific legal definition of hate speech. However, there are penal, civil and administrative provisions that protect citizens from malicious acts of discrimination associated with hate speech. To begin with, at the state level, the Spanish Constitution³³ is based on principles of democratic values and non-discrimination on the basis of race, sex, religion, opinion or any other personal or social condition or circumstance. Specifically, Part I sets forth the individual's fundamental rights and duties, such as human dignity (under Article 10) and equality before the law (under Article 14). In addition, the rights to honour, personal and family privacy and to one's image are guaranteed (under Article 18).

The Spanish Penal Code³⁴ is considered the primary legislation that prohibits criminal offences associated with hate speech such as threats (under Article 169), defamation – calumny (under Articles 205 to 207) and injury (under Articles 208–210) – and attacks on honour and moral dignity (under Article 173). Furthermore, committing these criminal offences against legally-protected groups of people is considered an aggravating factor (under Article 170), which results in increased penalties for the offenders. In 2015, following the ECRI General Policy Recommendation No. 7 (revised) on national legislation to combat racism and

³² Loi du 29 Juillet 1881 sur la liberté de la presse. Version consolidée au 08 Juillet 2014. https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=2D9E6AEE9BE04576DF46A63A4C088694.tpdjo05v_?cidTexte=LEGITEXT000006070722&dateTexte=20140708 (accessed 29 May 2020).

³³ Spanish Constitution 1978. <https://www.boe.es/legislacion/documentos/ConstitucionINGLES.pdf> Original title: Constitución Española 1978. BOE-A-1978.31229. [https://www.boe.es/eli/es/c/1978/12/27/\(1\)/con](https://www.boe.es/eli/es/c/1978/12/27/(1)/con) (accessed 30 May 2020).

³⁴ Organic Law No. 1/2015 of 30 March 2015, on Amendments to the Penal Code (Organic Law No. 10/1995 of 23 November 1995). http://melaproject.org/sites/default/files/2018-02/Spanish%20Criminal%20Code%20-%20Article%20510_0.pdf (accessed 30 May 2020).

racial discrimination,³⁵ the Spanish Penal Code was reformed. The most relevant improvement was that the essential elements of hate speech were included in the reformulation of Article 510, section 1, paragraph a) refers to the content, the purpose and the targets of hate speech:

Those who, directly or indirectly, foster, promote or incite hatred, hostility, discrimination or violence against a group, or part thereof, or against a certain person for belonging to such a group, for reasons of racism, anti-Semitism or for other reasons related to ideology, religion or beliefs, family circumstances, the fact that the members belong to an ethnicity, race or nation, national origin, gender, sexual orientation or identity, or due to gender, illness or disability (Criminal Code 2016: 214).³⁶

The reference made to hate speech in Article 510 is ambiguous and equivocal for various reasons. First, the speech acts constituting hate speech are presented as mutually exclusive options through the disjunctive conjunction “or” when in fact they may be concurrent (see Chapter 6 on complex speech acts) – e.g. “encourage”, “promote” or “incite”. Second, the various discourse strategies are presented as mutually exclusive when they may be not so – e.g. “directly” or “indirectly”. Third, the effects on the targets are too broadly defined, as “discrimination” is also added to “hatred”, “hostility” or “violence”. Fourth, the legally-protected characteristics are wide-ranging – e.g. “or other grounds relating to ideology, religion or belief, family status, membership of an ethnic group, race or nation, national origin, gender, sexual orientation or identity, illness or disability”. In paragraph b), Article 510 provides an exhaustive definition of dissemination. Nevertheless, the use of the disjunctive conjunction “or” hinders legal interpretation, as shown below:

Those who produce, prepare, possess with the purpose of distributing, provide third parties access to, distribute, publish or sell documents or any other type of material or medium that, due to the content thereof, are liable to directly or indirectly foster, promote or incite hatred, hostility, discrimination or violence against a group, or part thereof, or against a certain person for belonging to such a group, for reasons of racism, anti-Semitism or for other reasons related to ideology, religion or beliefs, family circumstances, the fact that the

³⁵ ECRI General Policy Recommendation No. 7 (revised) on National Legislation to Combat Racism and Rational Discrimination (adopted on 13 December 2002 and revised on 7 December 2017). <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.7> (accessed 31 May 2020).

³⁶ Criminal Code. 2016. Ministry of Justice. Technical Secretariat (eds.). Colección Traducciones del Derecho Español. Translated by Clinter Traducciones e Interpretaciones, S.A. Madrid. https://www.mjusticia.gob.es/es/AreaTematica/DocumentacionPublicaciones/Documents/Criminal_Code_2016.pdf.

members belong to an ethnicity, race or nation, national origin, gender, sexual orientation or identity, or due to gender, illness or disability [. . .] (Criminal Code 2016: 214).³⁷

Moreover, in paragraph c), Article 510 refers to other malicious acts relating to hate speech, such as publicly denying, seriously trivialising or glorifying the crime of genocide; crimes against humanity or crimes against persons and property protected in the event of armed conflict, or the glorification of the perpetrators of such crimes. In section 2, paragraph a), Article 510 prohibits damage to human dignity by any means of public expression and dissemination. Paragraph b) penalises exaltation or justification of offences committed by another group against the target groups by any means of public expression or dissemination.

Another crucial legislative reference is Law 19/2007 against Violence, Racism, Xenophobia and Intolerance in Sport,³⁸ which is also ambiguous and difficult for legal practitioners to interpret because it is too broad in terms of (a) the speech acts – e.g. it prohibits conduct intended to “threaten”, “intimidate”, “insult”, “humiliate” and “harass”; (b) the targets – e.g. individuals of vulnerable groups identified by legally-protected characteristics, such as “race”, “ethnic”, “geographical or social origin”, “religion”, “belief”, “disability”, or “sexual orientation”; and (c) the unlawful conduct -e.g. statements, gestures, singing, sounds, slogans and the display of banners, flags and symbols, amongst other instances.

6 Conclusions

In this chapter, Wittgenstein’s concept of family resemblance (Wittgenstein 2009 [1953]) has helped us understand that hate speech does not have a single meaning but a family of meanings. Ordinary language analysis showed the family of meanings the term hate speech has in non-technical, everyday language. The three legal scholarly approaches to defining hate speech – content-based, intent-based and harms-based – ended up in a family of legal scholar definitions of hate speech, as did the approaches to a technical legal definition of hate speech in international law, common law and civil law – European Union law and Member State law.

³⁷ Ibid, 214.

³⁸ Law 18/2007, Against Violence, Racism, Xenophobia and Intolerance in Sport. <https://www.global-regulation.com/translation/spain/1445414/law-19-2007%252c-of-july-11%252c-against-violence%252c-racism%252c-xenophobia-and-intolerance-in-sport.html> Original title: Ley 19/2007, de 11 de julio, contra la Violencia, el Racismo, la Xenofobia y la Intolerancia en el Deporte. <https://www.boe.es/buscar/pdf/2007/BOE-A-2007-13408-consolidado.pdf> (accessed 30 May 2020).

Throughout this chapter, I have highlighted eight elements that can help define hate speech rigorously. I concur with Sellars (2016) in that these elements are: (1) the targeting of a group that is identifiable by legally-protected characteristics; (2) the content of the message only expresses hatred; (3) the speaker intends to harm or to encourage harmful activity; (4) the speech causes harm; (5) the speech is likely to incite wrong actions beyond the speech itself; (6) the speech is delivered in public; (7) the context makes violent response possible due to, for example, the power of the speaker, the receptiveness of the audience and the history of violence in the area where the speech is delivered; and (8) the speech has no other redeeming purpose.

I conclude that hate speech is an umbrella term that refers to negative forms of conduct intended to publicly incite hatred and violence against groups identifiable by legally-protected characteristics. Hate speech, a type of negative social behaviour, is likely to harm the dignity and equality of the target groups, keep them marginalised from mainstream society and ultimately destroy social cohesion and peace.