



ALEXANDER BROWN*

WHAT IS HATE SPEECH? PART 1: THE MYTH OF HATE

(Accepted 28 February 2017)

ABSTRACT. The issue of hate speech has received significant attention from legal scholars and philosophers alike. But the vast majority of this attention has been focused on presenting and critically evaluating arguments for and against hate speech bans as opposed to the prior task of conceptually analysing the term ‘hate speech’ itself. This two-part article aims to put right that imbalance. It goes beyond legal texts and judgements and beyond the legal concept *hate speech* in an attempt to understand the general concept *hate speech*. And it does so using a range of well-known methods of conceptual analysis that are distinctive of analytic philosophy. One of its main aims is to explode the myth that emotions, feelings, or attitudes of hate or hatred are part of the essential nature of hate speech. It also argues that hate speech is best conceived as a family resemblances concept. One important implication is that when looking at the full range of ways of combating hate speech, including but not limited to the use of criminal law, there is every reason to embrace an understanding of hate speech as a heterogeneous collection of expressive phenomena. Another is that it would be unsound to reject hate speech laws on the premise that they are effectively in the business of criminalising emotions, feelings, or attitudes of hate or hatred.

I. INTRODUCTION

What does the state, acting on behalf of society as a whole, owe to citizens when it comes to regulating speech or other modes of expression? Some people believe that in answering this question it makes a positive difference whether or not the speech in question is insulting, degrading, defaming, negatively stereotyping or inciting hatred, discrimination or violence against people in virtue of their

* Alexander Brown is Senior Lecturer in Contemporary Social and Political Theory at the University of East Anglia (UEA). He joined UEA in 2009 prior to which he was a lecturer in legal and political theory at University College London (UCL) (2005–2009). He is the author of *Hate Speech Law: A Philosophical Examination* (Routledge, 2015), *Ronald Dworkin’s Theory of Equality: Domestic and Global Perspectives* (Palgrave, 2009), and *Personal Responsibility: Why it Matters* (Continuum, 2009).

race, ethnicity, nationality, religion, sexual orientation, disability, gender identity, for example; and that it makes a positive difference because such speech implicates issues of harm, dignity, security, healthy cultural dialogue, democracy, and legitimacy, to name just a handful of relevant issues. They sometimes use the term ‘hate speech’ to express that general view.¹ (In what follows I use italics for the concept *hate speech* and single speech marks for the term ‘hate speech’.) Critics of hate speech regulations, by contrast, claim that even though it has become ‘fashionable’ to defend such regulations and even though defenders of such regulations are (according to critics) ‘well-meaning’,² in reality hate speech regulations are themselves harmful to self-realization and autonomy, ineffective at best and often counterproductive, and damaging to democracy and legitimacy, amongst other things.³ Indeed, to defend hate speech regulations (critics maintain) belies an implicit wish to defend the regulation of speech that quite plainly should not be regulated, namely, merely offensive speech.⁴ Defenders of (some) hate speech regulations could, I suspect, respond that the critics are being patronising; that it is the critics who are adopting a fashionable and well-meaning yet wrongheaded position. Defenders could even retort that it is the critics who harbour a secret desire to deregulate all speech even speech that quite plainly should not be deregulated. But what exactly are the two sides arguing about here? What is hate speech? Until we can sensibly answer that question it seems that the

¹ See, e.g., Alexander Tsesis, *Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements* (New York, NY: New York University Press, 2002); Bhikhu Parekh, ‘Hate Speech: Is There a Case for Banning?’, *Public Policy Research* 12 (2005-2006): 213–223; Steven J. Heyman, *Free Speech and Human Dignity* (New Haven, CT: Yale University Press, 2008); Alexander Brown, *Hate Speech Law: A Philosophical Examination* (London: Routledge, 2015).

² See, e.g., Miklos Haraszti, ‘Forward’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012); Martha C. Nussbaum, ‘Law for Bad Behavior’, *The Indian Express*, February 21, 2014. Available at <http://indianexpress.com/article/opinion/columns/law-for-bad-behaviour/>.

³ See, e.g., C. Edwin Baker, ‘Hate Speech’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012); Ronald Dworkin, ‘Reply to Jeremy Waldron’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012); Robert Post, ‘Interview’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012); Nadine Strossen, ‘Interview’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012); Eric Heinze, *Hate Speech and Democratic Citizenship* (Oxford: Oxford University Press, 2016).

⁴ See, e.g., Robert Post, ‘Hate Speech’, in I. Hare and J. Weinstein (eds.) *Extreme Speech and Democracy* (Oxford: Oxford University Press, 2009); James Weinstein, ‘Hate Speech Bans, Democracy and Political Legitimacy’, *Constitutional Commentary*, forthcoming.

debate about hate speech regulations has no chance of moving forward, much less of being understood by bemused onlookers. Since every country on the planet has probably witnessed at some point in its history instances of what could be labelled 'hate speech', and virtually every legal system in the world contains at least one law that could be interpreted as a 'hate speech law', we are all of us caught up in the aforementioned debate whether we like it or not, and whether or not we know what the debate is actually about.

Numerous legal scholars have put forward putative definitions of the term 'hate speech'.⁵ Unsurprisingly, all of these legal scholars have sought, explicitly or implicitly, to define a legal concept *hate speech*, meaning a concept that refers to speech that is, or has been at one time or could be in the future, treated as hate speech for the

⁵ See, e.g., Mari Matsuda, 'Public Response to Racist Speech: Considering the Victim's Story', *Michigan Law Review* 87 (1989): 2320–2381, p. 2357; Suzanna Sherry, 'Speaking of Virtue: A Republican Approach to University Regulation of Hate Speech', *Minnesota Law Review* 75 (1991): 933–944, p. 933; Sandra Coliver, 'Hate Speech Laws: Do They Work?', in S. Coliver (ed.) *Striking a Balance: Hate Speech, Freedom of Expression and Non-Discrimination* (London: Article 19 International Centre Against Censorship/Human Rights Centre, University of Essex, 1992), p. 363; Frederick Schauer, 'Uncoupling Free Speech', *Columbia Law Review* 92 (1992): 1321–1357, p. 1349; Rodney Smolla, *Free Speech in an Open Society* (New York, NY: Vintage Books, 1992), p. 152; Charles Lawrence III et al., 'Introduction', in M. Matsuda et al. (eds.) *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Boulder, CO: Westview Press, 1993), p. 1; Ira Glasser, 'Introduction', in H. Gates et al. (eds.) *Speaking of Race, Speaking of Sex: Hate Speech, Civil Rights, and Civil Liberties* (New York, NY: New York University Press, 1994), p. 1; Laura Lederer and Richard Delgado, 'Introduction', in L. Lederer and R. Delgado (eds.) *The Price We Pay: The Case Against Racist Speech, Hate Propaganda, and Pornography* (New York, NY: Hill and Wang, 1995), pp. 4–5; Kathleen Sullivan and Gerald Gunther, *Constitutional Law* (New York, NY: Foundation Press, 1995), p. 1131; Larry Alexander, 'Banning Hate: Speech and the Sticks and Stones Defense', *Constitutional Commentary* 13 (1996): 71–100, p. 71; Bryan A. Garner, *Black's Law Dictionary, 7th Edn* (St. Paul, MN: West Group, 1999), pp. 1407–1408; John T. Nockleby, 'Hate Speech', in L. Levy and K. Karst (eds.) *Encyclopedia of the American Constitution, Vol. 3, 2nd Edn* (Detroit, MI: Macmillan, 2000), p. 1277; J. Angelo Corlett and Robert Franciscotti, 'Foundations of a Theory of Hate Speech', *The Wayne Law Review* 48 (2002): 1071–1100, p. 1083; Tsesis, *Destructive Messages*, p. 211, n. 1; Jon B. Gould, *Speak No Evil: The Triumph of Hate Speech Regulation* (Chicago, IL: Chicago University Press, 2005), p. 14; Raphael Cohen-Almagor, *The Scope of Tolerance: Studies on the Costs of Free Expression and Freedom of the Press* (London: Routledge, 2006), p. 153; Raphael Cohen-Almagor, 'Holocaust Denial is a Form of Hate Speech', *Amsterdam Law Forum* 2 (2009): 33–42, p. 35; Raphael Cohen-Almagor, 'Fighting Hate and Bigotry on the Internet', *Policy and Internet* 3 (2011): 1–26, pp. 1–2; James Weinstein and Ivan Hare, 'General Introduction: Free Speech, Democracy, and the Suppression of Extreme Speech Past and Present', in I. Hare and J. Weinstein (eds.) *Extreme Speech and Democracy* (Oxford: Oxford University Press, 2009), p. 4; Kathleen Mahoney, 'Hate Speech, Equality, and the State of Canadian Law', *Wake Forest Law Review* 44 (2009): 321–351, pp. 325–326; Post, 'Hate Speech', p. 127; Steven P. Lee, 'Hate Speech in the Marketplace of Ideas', in D. Golash (ed.) *Freedom of Expression in a Diverse World* (Dordrecht: Springer, 2010), p. 22; Douglas M. Fraleigh and Joseph S. Tuman, *Freedom of Expression in the Marketplace of Ideas* (Thousand Oaks, CA: Sage, 2011), p. 139; Susan Benesch, *Dangerous Speech: A Proposal to Prevent Group Violence* (New York: World Policy Institute Paper, 2012), p. 1; Eduardo Bertoni and Julio Rivera, 'The American Convention on Human Rights: Regulations on Hate Speech and Other Similar Expressions', in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012), p. 499; Kylie Weston-Scheuber, 'Gender and the Prohibition of Hate Speech', *Queensland University of Technology Law and Justice Journal* 12 (2012): 132–150, pp. 139–140.

purposes of legal regulation in one form or another. Some of these legal scholars seek to characterise a specific legal concept *hate speech* that relates to a particular body of law and legal regime, such as incitement to hatred laws in England and Wales. Others characterise a cluster legal concept *hate speech* that is associated with a class of hate speech laws, such as incitement to hatred laws. Yet others characterise an umbrella legal concept *hate speech* that implicates a range of different clusters or classes of hate speech law, such as laws proscribing group libel, media regulations limiting negative stereotyping or stigmatization, laws disallowing insult or denigration, laws banning incitement to hatred, and so on.⁶ Because these characterisations differ both in terms of whether they are grounded in what has been, what is, or what could be treated as hate speech within a body of law and in terms of the levels of generality at which they operate, they often wind up saying very different, often contradictory things about hate speech. Given that hate speech laws provoke such strong moral reactions, on the part of defenders and critics alike, and given that legal meanings will themselves draw on a range of deeper values and principles about which people reasonably disagree, it is no surprise that there remains such divergence over how to define the very term that stands at the epicentre of the disagreement.

However, I believe that in order to understand how best to respond to hate speech, whether this is via bodies of law and legal regimes and/or through a range of extralegal measures including counter-speech and education, we must pay serious attention to the fact that *hate speech* is not merely a legal concept, in the narrow sense of a concept that is used and in some cases even defined, explicitly or implicitly, within bodies of law and legal systems; it is also an ordinary concept that is (a) used by people who are not legal professionals or writers about the law, and (b) has a panoply of uses not only within bodies of law and legal systems but also within a range of other social, cultural, political and economic domains. If we focus too hard on a legal concept *hate speech* we risk prejudging what the right response to the problem of hate speech might be: for, we may have defined the nature and contours of the phenomena in question precisely with legal responses in mind. For defenders of hate speech laws, the danger is that if all one has is a hammer, everything starts

⁶ Cf. Brown, *Hate Speech Law*, ch. 2.

to look like a nail. Conversely, for critics of hate speech laws, if one is fixated on the idea that it is wrong to use a sledge hammer to crack a nut, then everything starts to look like a nut. Therefore, I position my analysis much closer to a different scholarly tradition, one that has attempted to analyse the meaning of the term 'hate speech' using various techniques of conceptual philosophical analysis.⁷ Nevertheless, my analysis also differs from that of these other philosophers in two main ways to be further explained and defended in the course of the articles: first, unlike them, I reject some of the implicit assumptions about hate speech carried over, consciously or unconsciously, from the work of legal scholars; second, unlike them, I argue that hate speech is a family resemblances concept that does not admit of definition.

Before setting out these points of differentiation, however, I need to clarify what I see as the ostensible differences between the legal concept (or class of concepts) *hate speech* and the ordinary concept *hate speech*. I do not mean to claim that the legal concept *hate speech* is somehow detached from the ordinary concept *hate speech* as though legislators, legal professionals and scholars of law operate separately from and with indifference to the rest of human life carrying on the background. Such a claim would be difficult to square with the fact that such people are all influenced or inspired by values and principles not merely of a legal character but also of a more general ethical, moral and political character, especially in areas like hate speech law, which concern what we owe to each other in matters of interpersonal interaction. Rather, what I mean to say is

⁷ See, e.g., Andrew Altman, 'Liberalism and Campus Hate Speech: A Philosophical Examination', *Ethics* 103 (1993): 302–317, pp. 309–310; Judith Butler, *Excitable Speech: A Politics of the Performative* (New York, NY: Routledge, 1997), p. 77; Susan Brison, 'The Autonomy Defense of Free Speech', *Ethics* 108 (1998): 312–339, p. 313; Susan Brison, 'Hate Speech', in H. La Follette (ed.) *The International Encyclopedia of Ethics* (Oxford: Wiley-Blackwell, 2013), p. 2332; David Brink, 'Millian Principles, Freedom of Expression, and Hate Speech', *Legal Theory* 7 (2001): 119–157, p. 119; L. W. Sumner, 'Hate Crimes, Literature and Speech', in R. Frey and C. Heath Wellman (eds.) *A Companion to Applied Ethics* (Oxford: Blackwell, 2003), p. 142; David Miller, *Political Philosophy: A Very Short Introduction* (Oxford: Oxford University Press, 2003), p. 65; Norman Fischer, 'First Amendment Morality Versus Civility Morality', in C. Sistare (ed.) *Civility and Its Discontents: Civic Virtue, Toleration, and Cultural Fragmentation* (Lawrence, KS: University of Kansas Press, 2004), p. 161; Parekh, 'Hate Speech', p. 214; Jeremy Waldron, 'Dignity and Defamation: The Visibility of Hate', *Harvard Law Review* 123 (2010): 1596–1657, pp. 1600–1601; Kevin M. Graham, *Beyond Redistribution: White Supremacy and Racial Justice* (Plymouth: Lexington Books, 2010), pp. 77–78; Caleb Yong, 'Does Freedom of Speech Include Hate Speech?', *Res Publica* 17 (2011): 385–403, p. 386; Rae Langton, 'Beyond Belief: Pragmatics in Hate Speech and Pornography', in I. Maitra and M. McGowan (eds.) *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University Press, 2012), pp. 74–77; Rae Langton et al., 'Language and Race', in G. Russell and D. Graff Fara (eds.) *Routledge Companion to the Philosophy of Language* (London: Routledge, 2012), pp. 757–760.

that how the term ‘hate speech’ is defined, explicitly or implicitly (via interpretation), within some body of law and legal system, cases and doctrines as well as statutes, is only the tip of the iceberg. Nowadays a great many different kinds of people who are not legislators, legal professionals or scholars of law use the term ‘hate speech’ in countless different types of context about a tremendous diversity of phenomena.

Now it is certainly true that the term ‘hate speech’ was coined by a group of legal scholars in the late 1980s in the United States (more on this in ‘What is Hate Speech? Part 2: Family Resemblances’, henceforth ‘Family Resemblances’) in response to what they saw as the way different legal systems tackled certain sorts of harmful racist speech. When Mari Matsuda first used the term ‘hate speech’ in her seminal article from 1989, ‘Public Response to Racist Speech: Considering the Victim’s Story’, her central purpose in using the term ‘hate speech’ was to highlight the way in which the legal system in the United States failed victims of harmful racist speech by providing them with inadequate means of seeking redress, civil or criminal. In the article, she includes a number of legal cases and also some examples not associated with actual legal proceedings and not easily actionable under current laws. But these examples were intended to show the limitations of the legal concept *hate speech* in the United States, as a form of *protected speech*. She contrasted the American-based legal concept *hate speech* with an ideal type or model version of what the legal concept *hate speech* could and should be. Perhaps in the process she was also invoking specific legal concepts *hate speech* found in Europe. Because of the work of legal scholars like Matsuda, the term ‘hate speech’ has now been taken up by legislators and legal professionals themselves more explicitly. However, and this is the crucial part, this new bit of terminology or legalese has now also been taken up by the media and by ordinary people and has taken on a life of its own.

By way of evidence for this last point, entering the term ‘hate speech’ into Google returns more than eighty million results.⁸ No doubt a significant proportion of these results are about the concept *hate speech* as it pertains to hate speech law. Indeed, it seems likely that the term ‘hate speech’ (and its non-English equivalents) often

⁸ On December 8, 2016.

first comes to the attention of ordinary people through television and radio programmes, newspaper and magazine articles, Internet message boards, social networking websites, and Internet news channels that mention the issue of hate speech in relation to high profile court cases, especially those involving already recognised public figures. Consider well-reported hate speech cases involving the French politician Brigitte Bardot, the Australian journalist Andrew Bolt, the Indian politician Akbaruddin Owaisi, the French comedian Dieudonné M'bala M'bala, the Kenyan politicians Moses Kuria and Junet Mohamed, and the Dutch politician Geert Wilders. These and similar cases around the world have all become trending stories on social media and online news channels. Nevertheless, the term 'hate speech' is also used by people who are not legal professionals and/or writers about the law, and in the course of talking about events that are not connected with legal proceedings and where there is virtually no possibility of legal proceedings. So, for example, at the time of writing, entering the search term 'hate speech Donald Trump' into Google returned over four million results. Indeed, a research group funded by the European Union has recently undertaken a survey of the media across Europe using the European Media Monitor (EMM) tool to find hits of the term 'hate speech'. The preliminary figures for the UK shows an average of just under 100 hits per month between April 2015 and February 2016, with a spike of 224 hits in December 2015 with many hits occurring within stories about the campaign speeches of Donald Trump.⁹

Hitherto, the term 'hate speech' has been perhaps most often associated with liberal progressives, or people on the left of politics – who use it to highlight and problematise speech that they view as racist, xenophobic, homophobic, Islamophobic, misogynistic, disablist, or in some other way targeted at minority groups in ways that supposedly violate ideals of respect, solidarity, tolerance, and so forth. By contrast, many political and religious conservatives repudiate such uses of the term, and view them simply as crude attempts to close down meaningful debate on what they believe are the evils of open-border policies, the failures of multiculturalism as a social experiment, the lamentable decline of traditional moral values, political correctness gone mad, and so on. Yet even people who are

⁹ Preliminary figures obtained from Professor Sharon Millar, University of Southern Denmark, a member of the European Union funded research network C.O.N.T.A.C.T.

deeply critical of what they see as an inordinate and dangerous obsession with identity-based speech will use the term ‘so-called hate speech’, thereby implicitly acknowledging that the term ‘hate speech’ has an ordinary meaning – albeit they disapprove of this term given its ordinary meaning. More importantly, some political and religious conservatives have now consciously adopted terms like ‘hate speech’, ‘hate mail’, and ‘hate’ to describe speech on the part of liberal progressives and civil rights activists – speech which, in the eyes of those political or religious conservatives, picks them out and attacks them on the basis of their conservative beliefs.¹⁰

In addition to this, the ordinary concept *hate speech* seems to be playing an increasing role in popular culture. Consider the *South Park* episode from 2005 in which Cartman’s school presentation about ‘ginger kids’ – who it is claimed are suffering from ‘gingervitis’ and have ‘no souls’ – causes Kyle to object, ‘That’s not a lecture, it’s a hate speech.’¹¹ Or take the furore surrounding the singer Beyoncé’s halftime performance at Super Bowl 50 in 2016. Some people – including members of the All Lives Matter social movement – took to social media to describe the fact that Beyoncé’s backing dancers were dressed like Black Panther members, along with the content of her recent music videos and public statements in support of the Black Lives Matter movement, as ‘hate speech’.¹² And no doubt there were defenders of Beyoncé’s public stance on the injustices faced by people of colour in the United States who viewed the portrayal of Beyoncé’s performance and music videos as ‘hate speech’ as itself hate speech. Or bear in mind what the Harry Potter author, J. K. Rowling tweeted after it emerged that Donald Trump won the 2016 presidential election.

We stand together. We stick up for the vulnerable. We challenge bigots. We don’t let hate speech become normalized. We hold the line.¹³

¹⁰ See, e.g., interviews with, and sermons given by, the United States evangelical Christian pastor Steven Anderson broadcast as part of the television documentary *Preachers of Hate*, aired on the BBC UK, October 11, 2016.

¹¹ *South Park*, Season 9, Episode 11, aired on *Comedy Central* US, November 9, 2005.

¹² Adam Howard, ‘Beyoncé Super Bowl Performance Inspires Protest, Counter-protest’, *NBC News Online*, February 10, 2016. Available at <http://www.nbcnews.com/news/nbcblk/beyonc-super-bowl-performance-inspires-protest-counter-protest-n515996>.

¹³ J. K. Rowling, @jk_rowling, Twitter, November 8, 2016. Available at https://twitter.com/jk_rowling/status/796252371739430913.

There is always a chance, of course, that as society as a whole co-opts the term 'hate speech' and puts it to myriad purposes, the term itself will be emptied of its original, legal-technocratic meaning to such an extent that it becomes merely an empty vessel; a generic term of disapproval that has the same semantic content as, 'I really don't like what you said'. But even if we are on the path toward emptying the term 'hate speech' of substantive meaning, I do not think we are there yet. The point is that the term 'hate speech' is used because it is useful, and it will remain useful so long as it can be used to do more than merely signal disapproval. If all it did were signal disapproval it would soon fall out of fashion or be replaced by newer, cooler bits of language that did the same thing but in more interesting ways.

My hunch is that many ordinary people have been willing to use the term 'hate speech' (and its non-English equivalents) for much the same reason that legal scholars coined the term in the first place, namely, that it provides a rough but nevertheless serviceable term to describe phenomena that have been the subject of legal sanctions, of one kind or another, since the Roman laws on group defamation¹⁴ but that, in all likelihood, have also been present throughout human history beginning with the earliest multiethnic societies of the ancient world,¹⁵ that is, the expressive dimensions of identity-based envy, hostility, conflict, mistrust and oppression. It is also worth remembering at this point that the term 'hate speech' is only the latest in a succession of terms that historically have been used to refer to forms of speech that attack members of groups or classes people identified by certain protected characteristics, including, for instance, 'race hate', 'group libel' and 'hate propaganda'.¹⁶ So for some time societies have found good use for this family of terms. But what is immediately noticeable, I think, about the term 'hate speech' is its generality, the way it can be used to meaningfully talk about far more protected characteristics and far more varieties of speech than any of its predecessors. Thus it seems to me that it makes a material difference that people use the term 'hate speech' rather than these

¹⁴ See, e.g., David Riesman, 'Democracy and Defamation: Control of Group Libel', *Columbia Law Review* 42 (1942): 727–780, p. 728.

¹⁵ See, e.g., Lionel Casson, *Everyday Life in Ancient Egypt, Revised and Expanded Edition* (Baltimore: The Johns Hopkins University Press, 2001), pp. 58–59.

¹⁶ See Samuel Walker, *Hate Speech: The History of an American Controversy* (Lincoln, NE: University of Nebraska Press, 1994), p. 8.

older terms, and it makes a difference because the term is referring to a much broader and more capacious concept; and this is not necessarily a bad thing.

Judith Butler once wrote, ‘the category [hate speech] cannot exist without the state’s ratification’.¹⁷ As a claim about the legal concept *hate speech* this is trivially true. But there is surely an ordinary concept *hate speech* whose existence is not dependent on the state’s ratification. Butler does not explain in unambiguous detail what she means by ‘ratification’, but suppose it means that when the state takes on the role of deciding what may or may not be permissibly said, it in some sense *produces* legal categories, such as the legal category *hate speech*. (This means that even a state which assumes the aforementioned role but decides not to ban hate speech nevertheless creates the category *hate speech*, which goes into the permissible or protected speech column.¹⁸) This point about the dependence of the category on the state’s ratification is much harder to sustain in the case of the ordinary concept *hate speech*, however. Whether or not the state takes on the role of line-drawing in relation to acceptable and unacceptable speech, there may well remain an ordinary concept *hate speech*. I do not mean to say that the state’s decision in this regard has no impact on the content of the ordinary concept *hate speech*. The ordinary concept is constituted by how ordinary people use the term ‘hate speech’ and this is likely to be influenced by how legal professionals and scholars of law use the term ‘hate speech’ and vice versa. So if the state decided to refrain from line-drawing, the ordinary concept *hate speech* might not exist in exactly the same form. But this concept would persist, I think, precisely because it reflects ordinary people’s sense of what is acceptable and unacceptable speech in relation to groups or classes of persons identified by protected characteristics, based on ordinary people’s moral values and principles. Now it is also true that the ordinary concept *hate speech* probably includes the idea that hate

¹⁷ Butler, *Excitable Speech*, p. 77.

¹⁸ I am not sure how Butler thinks that the state could avoid producing the category *hate speech* given that the state does not operate in a vacuum and once groups come forward to request protection against hate speech, the state will make a determination about whether or not to meet the request, and either way it will make some sort of category *hate speech*, either as regulatable speech or protected speech. Suppose public order law was radically reformed in England and Wales, with all the offences that the legislators and legal scholars interpret as ‘hate speech’ offences being repealed, thus going the same way as the blasphemy offence in England and Wales. In that case it would still be producing a category *hate speech*, only this time as speech that is not to be treated as a public order offence.

speech is *regulatable speech* (in the sense that there is a *pro tanto* reason to regulate it). But at least as regards to the ordinary concept this feature of the concept does not depend on the state's ratification. Ordinary users of the term 'hate speech' could think that hate speech is regulatable speech even if the state itself decided to withdraw from line-drawing in relation to acceptable and unacceptable speech and/or there was no immediate prospect of hate speech actually being regulated.

Interestingly, there is precedent for drawing a distinction between the legal concept and the ordinary concept *hate speech* within normal legal practice itself. It is customary for judges to mark the distinction between the legal meaning of key words in statutes and the ordinary meaning of those words (e.g., the word 'insulting' discussed in *Brutus v. Cozens*¹⁹), whether it is to argue that the legal meaning need not depend on the ordinary meaning or to insist instead that other things remaining equal the legal meaning should be interpreted primarily according to the ordinary meaning (known as the ordinary meaning principle or plain meaning rule).²⁰ I wish to invoke something like this ubiquitous distinction within legal interpretation here; albeit, perhaps unordinarily, I am interested in a term ('hate speech') whose ordinary meaning has arguably developed in the light of, and certainly after, the birth of the legal meaning.

How does one get at the ordinary meaning? Justice Scalia once wrote about the idea of ordinary meaning in legal interpretation that 'the acid test of whether a word can reasonably bear a particular meaning is whether you could use the word in that sense at a cocktail party without having people look at you funny.'²¹ I suspect that in the late 1980s and early 1990s the only cocktail parties in which the term 'hate speech' could have been used without people being looked at funny would have been those attended by legal scholars. But things are different now, the term 'hate speech' has entered into the fields of politics, popular culture, the arts, sport, media, education, science, and many more besides. So the cocktail party test must involve a party with the most diverse set of guests imaginable. Of course, the guests will also include legislators and

¹⁹ [1973] AC 854.

²⁰ See Brian G. Slocum, *Ordinary Meaning A Theory of the Most Fundamental Principle of Legal Interpretation* (Chicago, IL: University of Chicago Press, 2015).

²¹ *Johnson v. United States* 529 U.S. 694 (2000), Justice Scalia's dissenting opinion.

legal professionals, since they are not merely influenced by ordinary meanings but will inevitably influence ordinary meanings by adopting and promulgating certain legal meanings. At any rate, the basic idea of the cocktail party is that ordinary meaning has something to do with linguistic intuitions, part of the stock-in-trade of analytical philosophers. I shall say more about this and other aspects of my methodology in Section II.

In saying that more needs to be done to excavate the ordinary concept *hate speech* – so as to augment and, if necessary, offset the work that has already been done on the legal concept *hate speech* – I am advocating a new research agenda, to be pursued from a variety of academic disciplines across the arts, humanities, and social sciences. All too frequently when academics from other disciplines have turned their attention to the topic they have followed legal scholars in focusing on the legal concept *hate speech*. For example, as I have already mentioned, over the past twenty-five years or so a number of generalist philosophers have attributed meanings to the term ‘hate speech’,²² but more often than not these philosophers have not strayed very far from that which is prohibited by campus hate speech codes, certain types of municipal ordinances, and incitement to hatred laws.

What is particularly striking to me about the philosophical literature is the extent to which it echoes the work of legal scholars. I mention five notable overlaps. First, when legal scholars working in the field of critical race theory identified hate speech with various ways of subordinating, wounding, humiliating, or degrading historically oppressed or victimised racial groups,²³ arguably they laid the groundwork for those philosophers who would later utilise speech act theory to understand hate speech,²⁴ albeit these critical race theorists and these philosophers of language were themselves influenced by the work of Catharine MacKinnon and Andrea

²² See note 7 above.

²³ See, e.g., Matsuda, ‘Public Response to Racist Speech’, pp. 2358, 2363; Charles Lawrence III, ‘Cross Burning and the Sound of Silence: Anti-Subordination Theory and the First Amendment’, *Villanova Law Review* 37 (1992): 787–804, p. 792; Lawrence et al., ‘Introduction’, p. 1.

²⁴ See, e.g., Altman, ‘Liberalism and Campus Hate Speech’, pp. 309–310; Langton, ‘Beyond Belief’, pp. 74–77; Langton et al., ‘Language and Race’, pp. 757–760; Ishani Maitra, ‘Subordinating Speech’, in I. Maitra and M. McGowan (eds.) *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University Press, 2012), pp. 96–118; Mary Kate McGowan, ‘On ‘Whites Only’ Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination’, in I. Maitra and M. McGowan (eds.) *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University Press, 2012), p. 122; Brown, *Hate Speech Law*, pp. 75–86.

Dworkin, who originally identified pornography as subordinating speech in the context of proposing anti-pornography laws.²⁵ Second, the legal scholar Kent Greenawalt analyses the term 'hate speech' by appealing to the legal doctrines of fighting words, hostile environment harassment, and group libel,²⁶ and much the same characterisation can be found in the work of the philosophers Susan Brison and Timothy C. Shiell.²⁷ Third, several legal scholars have drawn distinctions between immediate, instant, face-to-face, targeted, and individualised hate speech, on the one hand, and indirect, diffuse, generalised, and impersonal hate speech, on the other hand,²⁸ and these same distinctions are replicated by the philosopher Caleb Yong.²⁹ Fourth, numerous legal scholars have implied that emotions, feelings, or attitudes of hate or hatred are part of the essential nature of hate speech and what distinguishes such speech from other kinds of speech and that the meaning of the term 'hate speech' is compositional, that is, a function of the literal meanings of the terms 'hate' and 'speech',³⁰ and similar assumptions can be found in the work of some philosophers, notably Norman Fischer and Bhikhu Parekh.³¹ Finally, scholars of hate speech law often proceed on the assumption that the term 'hate speech' is univocal and definable (it

²⁵ See, e.g., Catharine MacKinnon, *Feminism Unmodified* (Cambridge: Cambridge University Press, 1987); Catharine MacKinnon, *Only Words* (Cambridge, MA: Harvard University Press, 1993); Catharine MacKinnon and Andrea Dworkin (eds.) *In Harm's Way: The Pornography Civil Rights Hearings* (Cambridge, MA: Harvard University Press, 1997).

²⁶ Kent Greenawalt, *Fighting Words: Individuals, Communities and Liberties of Speech* (Princeton, NJ: Princeton University Press, 1995), chs. 4 and 5.

²⁷ Brison, 'The Autonomy Defense of Free Speech', p. 314; Timothy C. Shiell, *Campus Hate Speech on Trial, 2nd Edn* (Lawrence, KS: University of Kansas Press, 2009), p. 162.

²⁸ See, e.g., Alan E. Brownstein, 'Hate Speech and Harassment: The Constitutionality of Campus Codes that Prohibit Racial Insults', *William and Mary Bill of Rights Journal* 3 (1994): 179–217, p. 179; Greenawalt, *Fighting Words*, p. 63; Nicholas Wolfson, *Hate Speech, Sex Speech, Free Speech* (Westport, CT: Praeger Publishers, 1997), p. 60; Steven H. Shiffrin, *Dissent, Injustice, and the Meanings of America*. Princeton, NJ: Princeton University Press, 1999), pp. 76–77; Thomas Peard, 'Regulating Racist Speech on Campus', in C. Sistare (ed.) *Civility and Its Discontents: Civic Virtue, Toleration, and Cultural Fragmentation* (Lawrence, KS: University of Kansas Press, 2004), p. 142; Richard Delgado and Jean Stefancic, 'Four Observations About Hate Speech', *Wake Forest Law Review* 44 (2009): 353–370, pp. 362–363; Heinze, *Hate Speech and Democratic Citizenship*, p. 282.

²⁹ Yong, 'Does Freedom of Speech Include Hate Speech?', pp. 394–396.

³⁰ See, e.g., Glasser, 'Introduction', p. 1; Garner, *Black's Law Dictionary*, pp. 1407–1408; Nadine Strossen, 'Incitement to Hatred: Should There Be a Limit?', *Southern Illinois University Law Journal* 25 (2001): 243–279, p. 244; Corlett and Francescotti, 'Foundations of a Theory of Hate Speech', p. 1083; Post, 'Hate Speech', pp. 123–125; Weinstein and Hare, 'General Introduction', p. 4; Iginio Gagliardone et al., *Mapping and Analysing Hate Speech Online: Opportunities and Challenges for Ethiopia* (Oxford: University of Oxford Programme in Comparative Media and Law Policy, 2014), p. 9.

³¹ Fischer, 'First Amendment Morality Versus Civility Morality', pp. 161–162; Parekh, 'Hate Speech', p. 214.

has a single meaning, even if complex or disjunctive, and this meaning can be defined),³² and some philosophers seem to labour under the same assumption.³³ Associated with this particular assumption is a tendency to be oblivious to important differences between the writer's proffered definition and those of other scholars, and even to differences between the writer's own definitions from earlier to later work, as though the differences are inconsequential or that they along with all other writers are essentially defining the term in the same way only using slightly different language (language that can be used interchangeably).

I find some of these overlaps and shared understandings among generalist philosophers and legal scholars entirely unproblematic. The first three overlaps, I think, are extremely plausible and perhaps illuminate something of the subtle interplay and reciprocal relationship between the ordinary and legal concepts of *hate speech*. Nevertheless, in this article I do also want to challenge *some* of this received wisdom: specifically, the putative wisdom that emotions, feelings, or attitudes of hate or hatred are part of the essential nature of hate speech, what I call *the myth of hate*; and the conventional idea that the term 'hate speech' is univocal or has a single meaning, call this *the myth of compositionality*. Now I do not claim, nor should we believe anyone who did try to claim, that challenging this received wisdom is likely to be decisive in settling disputes about whether hate speech laws are warranted. These disputes are far too deep-rooted in our understanding of a range of values and principles to be easily resolved. But challenging this received wisdom could, I think, at least remove some of the obstacles to progress, such as if, say, one side of the debate is doing more reasoning than the other on the strength of false impressions about what hate speech is. I shall try to draw out these implications at the end of 'Family Resemblances'.

Before I begin, however, let me be clear that by focusing on concepts and words I do not mean to overlook the fact that what is at stake here are interpersonal phenomena that involve the interaction of actual human beings and often very real harms. On the

³² See, e.g., Cohen-Almagor, *The Scope of Tolerance*, p. 153; Cohen-Almagor, 'Holocaust Denial is a Form of Hate Speech', p. 35; Cohen-Almagor, 'Fighting Hate and Bigotry on the Internet', pp. 1–2.

³³ See, e.g., Altman, 'Liberalism and Campus Hate Speech', pp. 309–310; Brison, 'The Autonomy Defense of Free Speech', p. 313; Brison, 'Hate Speech', p. 2332; Brink, 'Millian Principles, Freedom of Expression, and Hate Speech', p. 119.

contrary, the point of focusing on concepts and terms is to gain access to or illuminate our understanding of the phenomena in question. In the words of Austin, '[w]hen we examine what we should say when, what words we should use in what situations, we are looking again not *merely* at words (or 'meanings', whatever they may be) but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena.'³⁴

The remainder of the first part of this article is structured as follows. Section II outlines existing attempts to characterise the legal concept *hate speech* and highlights some of the inconsistencies and contradictions in this existing body of characterisations. Section III examines the ordinary concept *hate speech* and assumes for the sake of argument that it is a compositional concept: that the meaning of the whole concept is a function of the meanings of the parts that make up the whole. Focusing on the assumption that emotions, feelings, or attitudes of hate or hatred are one basic building block of the ordinary concept *hate speech*, I proceed to argue that despite the existence of several possible connections between speech and hate or hatred none appear to capture an essential feature or defining quality of hate speech. This in turn undermines the compositionality assumption. Section IV looks at two, more nuanced versions of the idea that the term 'hate speech' has compositional semantics: the first, that it is a semi- or quasi-compositional term akin to 'zebra crossing' or 'pillow talk'; the second that it is part of a set of core-dependent homonyms. Although these approaches have some plausibility, I argue that they only serve to underline the speciousness of the myth of hate.

II. THE LEGAL CONCEPT(S) *HATE SPEECH*

In this section, I shall use the method of *conceptual jurisprudence* to analyse the legal concept(s) *hate speech*. This method involves explicating concepts that emerge from given bodies of law and legal practices. This includes looking at the way the term 'hate speech' has been defined in different human rights instruments, statutes, and legal cases, both positively ('hate speech is...') and negatively ('hate speech is not...'). But it also includes an attempt to reconstruct the

³⁴ J. L. Austin, 'A Plea for Excuses: The Presidential Address', *Proceedings of the Aristotelian Society* 57 (1956-7): 1-30, p. 8.

legal concept(s) *hate speech* from the occurrences and meanings of various linked or associated concept-terms that appear in the relevant bodies of law and legal practices – terms such as ‘group defamation’, ‘incitement to hatred’, ‘the circulation of ideas based on inferiority’, ‘racist propaganda’, ‘speech based on xenophobia, homophobia, Islamophobia, and anti-Semitism’, ‘group vilification’, ‘violation of dignity’, ‘discriminatory harassment’, ‘racist fighting words’, ‘Holocaust denial’, and so on. Part of the aim of conceptual jurisprudence is to capture the fundamental quality of the legal concept under investigation as well as something of its relationship to other legal concepts, principles and doctrines in the relevant bodies of law and legal practices.

I start by observing that the original uses of the then neologism ‘hate speech’ belonged to legal academics and journalists writing in academic journals and books and mainstream newspapers and magazines. They tended to use the term ‘hate speech’ for the purposes of classifying and make generalisations about a particular group of laws. For example, the Oxford English Dictionary entry for the term ‘hate speech’ offers only three illustrations of uses.³⁵ The first comes from a *Newsweek* article from 1988 in which Tamar Jacoby uses the term ‘hate speech’ in discussing the merits of university regulations designed to curb discriminatory harassment on university campuses (campus speech codes).³⁶ The second is from Samuel Walker’s 1994 book *Hate Speech: The History of an American Controversy* in which the historian contrasts the constitutional protection of ‘offensive forms of expression’ in the United States with the fact that ‘[a]lmost every [other] country prohibits hate speech directed at racial, religious, or ethnic groups’.³⁷ The third is from an article appearing in *Business Day*, a national daily newspaper in South Africa, from 2001 outlining aspects of the Promotion of Equality and Prevention of Unfair Discrimination Act 2000 which, amongst other things, prohibits hate speech.³⁸

Yet the term ‘hate speech’ is certainly not the preserve of legal scholars. It is now being utilised in the decisions of some interna-

³⁵ Oxford English Dictionary, ‘Hate Speech’, July 2002. Available at www.oed.com.

³⁶ Tamar Jacoby, ‘Time to Outlaw Racial Slurs? New Questions About Protecting Hateful Speech’, *Newsweek*, June 6, 1988.

³⁷ Walker, *Hate Speech*, p. 1.

³⁸ Unkown, *Business Day*, January 28, 2001, p. 13.

tional human rights courts and in some domestic legislation. In *Surek v. Turkey*³⁹ and *Gündüz v. Turkey*,⁴⁰ for example, the European Court of Human Rights (ECtHR) directly and explicitly employed a definition of the term ‘hate speech’ that had been given by the Committee of Ministers of the Council of Europe back in 1997: ‘the term ‘hate speech’ shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism or other forms of hatred based on intolerance [...]’.⁴¹ In both cases the Court also reiterated that to count as hate speech it is not sufficient that the words merely offend, shock or disturb; they must be capable of inciting hatred or violence. Or, to give one example from domestic law, in South Africa s. 10 of the Promotion of Equality and Prevention of Unfair Discrimination Act 2000 outlines a set of offences under the heading ‘Prohibition of Hate Speech’, including most notably s. 10(1)(c), ‘[...] no person may publish, propagate, advocate or communicate words based on one or more of the prohibited grounds, against any person, that could reasonably be construed to demonstrate a clear intention to: (a) be hurtful; (b) be harmful or to incite harm; (c) promote or propagate hatred.’ Likewise, s. 16(2) of the South African Bill of Rights specifies that the right to freedom of expression does not extend to ‘advocacy of hatred that is based on race, ethnicity, gender or religion, and that constitutes incitement to cause harm’.

A variant of the current approach also assumes that hate speech is primarily a legal concept but does not limit the materials of analysis to explicit attempts to clarify the concept within international and domestic law. In other words, it seeks to analyse the legal concept(s) *hate speech* through bodies of law that may not contain the exact term ‘hate speech’. It does so by picking out certain similarities or common traits that could be interpreted as elements of the legal concept *hate speech*. How on this approach do we know which laws and legal cases are the correct ones to look at? The ECtHR’s Press Unit points in one useful direction. Its 2016 ‘Factsheet – Hate Speech’ does not attempt to lay down a universally accepted definition of the

³⁹ No. 26682/95, Strasbourg, July 8, 1999. At para. 62, and Partly Dissenting Opinion of Judge Palm.

⁴⁰ No. 35071/97, Strasbourg, June 14, 2004. At paras. 21–22, 40–1, 44, 48, 51, and Dissenting Opinion of Judge Türmen.

⁴¹ Recommendation No. R (97) 20, October 30, 1997, Appendix. Available at <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680505d5b>.

term ‘hate speech’ but instead points to various conventions and a host of legal cases which in its view pertain to hate speech.⁴² Following this lead, progress could be made by searching laws and legal cases for words that are indicative of, contiguous with, or being used as proxies for the term ‘hate speech’ – for example, ‘hate’, ‘hatred’, ‘contempt’, ‘hostility’, ‘enmity’, ‘inferiority’, ‘racist propaganda’, ‘xenophobia’, ‘anti-Semitism’, ‘aggressive nationalism’, ‘homophobia’, ‘Islamophobia’, ‘antisemitism’, ‘group defamation’, ‘group vilification’, ‘insult’, ‘negative stereotyping’, ‘stigmatisation’, ‘humiliation’, ‘degradation’, ‘violation of dignity’, ‘discriminatory harassment’, ‘intolerance’. I shall not attempt here to provide a complete inventory of all clusters of laws that employ or exhibit this terminology and that could be used as source material for an interpretation of the legal concept(s) *hate speech*.⁴³ It is enough for my present purposes to indicate what this sort of approach looks like, and to outline some of the results it has yielded.

A small army of legal scholars and lexicographers have set forth characterisations of hate speech using the aforementioned approach. These characterisations exhibit numerous points of consensus but are also marked by significant inconsistencies and contradictions. First, consider how they deal with the connection between hate speech and hate or hatred. Some characterisations take hate speech literally to mean speech that expresses or communicates emotions, feelings, or attitudes of hate or hatred,⁴⁴ others imply that the motive of hatred is almost all there is to hate speech,⁴⁵ and still others identify hate speech as that which promotes, incites, or stirs up hatred.⁴⁶

Second, there is the issue of identifying which groups are covered by the legal concept *hate speech*. While most characterisations pro-

⁴² ECtHR Press Unit, *Factsheet – Hate Speech*, June, 2016. Available at www.echr.coe.int/Documents/FS_Hate_speech_ENG.pdf.

⁴³ Cf. Brown, *Hate Speech Law*, ch. 2.

⁴⁴ See, e.g., Garner, *Black’s Law Dictionary*, pp. 1407–1408; Strossen, ‘Incitement to Hatred’, p. 244; Oxford English Dictionary, ‘Hate Speech’; Corlett and Francescotti, ‘Foundations of a Theory of Hate Speech’, p. 1083; Post, ‘Hate Speech’, p. 123; Weinstein and Hare, ‘General Introduction’, p. 4; USLegal.com, ‘Hate speech’, June 2010. Available at <http://definitions.uslegal.com/h/hate-speech/>.

⁴⁵ See, e.g., Glasser, ‘Introduction’, p. 1.

⁴⁶ See, e.g., Cohen-Almagor, *The Scope of Tolerance*, p. 11; Cohen-Almagor, ‘Holocaust Denial is a Form of Hate Speech’, p. 35; Weston-Scheuber, ‘Gender and the Prohibition of Hate Speech’, pp. 139–140.

vide a list of group-identifying characteristics, including but not limited to race, ethnicity, religion, nationality, gender or gender identity, sexual orientation, disability,⁴⁷ some specify not a list of characteristics but a single core feature – for example, that the speech in question is about or against historically oppressed or victimised groups.⁴⁸ Some make no mention of the kinds of groups that are germane to the legal concept *hate speech*,⁴⁹ which is surprising given that almost without exception all actual hate speech laws attempt to specify which groups fall within the scope of the law.

Third, there is the question of the nature of the speech. Whereas some characterisations depict hate speech as a special case of insulting speech (i.e., insulting about or insulting to members of certain groups)⁵⁰ or offensive speech (i.e., offensive about or offensive to members of certain groups),⁵¹ other characterisations describe hate speech as ‘hostile verbal abuse’,⁵² ‘malicious speech’,⁵³ ‘speech that denigrates people’,⁵⁴ ‘speech that puts people down’,⁵⁵ or ‘antisocial oratory’.⁵⁶ Moreover, only some characterisations stress that hate speech, properly called, typically causes or has a substantial likelihood of causing harm of one kind or another.⁵⁷

Fourth, most characterisations make it clear that hate speech is not simply speech but is also performative or can involve speech acts of various types. Yet even here there are important differences. Some describe hate speech in such broad terms as to mean that a large

⁴⁷ See, e.g., Human Rights Watch, *‘Hate Speech’ and Freedom of Expression: A Human Rights Watch Policy Paper* (New York, NY: Fund for Free Expression, 1992), p. 1; Tsesis, *Destructive Messages*, p. 211, n. 1; Richard Delgado and Jean Stefancic, *Understanding Words That Wound* (Boulder, CO: Westview Press, 2004), p. 11; Cohen-Almagor, ‘Fighting Hate and Bigotry on the Internet’, pp. 1–2; Bertoni and Rivera, ‘The American Convention on Human Rights’, p. 499.

⁴⁸ See, e.g., Matsuda, ‘Public Response to Racist Speech’, p. 2357; Walker, *Hate Speech*, pp. 8–9; Cohen-Almagor, *The Scope of Tolerance*, p. 153.

⁴⁹ See, e.g., Kermit L. Hall, ‘Hate Speech’, in K. L. Hall (ed.) *The Oxford Companion to the Supreme Court of the United States, 2nd Edn* (Oxford: Oxford University Press, 2005), p. 426.

⁵⁰ See, e.g., Alexander, ‘Banning Hate’, p. 71; Larry Alexander, ‘Hate Speech’, in R. Newman (ed.) *The Constitution and Its Amendments, Vol. 3* (New York, NY: Macmillan Reference, 1998), p. 14.

⁵¹ See, e.g., Human Rights Watch, *‘Hate Speech’ and Freedom of Expression*, p. 1, n. 1; Sullivan and Gunther, *Constitutional Law*, p. 1131; Ivan Hare, ‘Hate speech’, in P. Cane and J. Conaghan (eds.) *The New Oxford Companion to Law* (Oxford: Oxford University Press, 2008), p. 520; Post, ‘Hate Speech’, p. 12.

⁵² Oxford English Dictionary, ‘Hate Speech’.

⁵³ Cohen-Almagor, ‘Fighting Hate and Bigotry on the Internet’, p. 1.

⁵⁴ Benesch, *Dangerous Speech*, p. 1.

⁵⁵ Anthony Cortese, *Opposing Hate Speech* (Westport, CT: Praeger Publishers, 2006), p. 1.

⁵⁶ Tsesis, *Destructive Messages*, p. 211, n. 1.

⁵⁷ See, e.g., Richard Delgado, ‘Toward a Legal Realist View of the First Amendment’, *Harvard Law Review* 113 (2000): 778–802, pp. 787–788; Tsesis, *Destructive Messages*, p. 211, n. 1; Delgado and Stefancic, *Understanding Words That Wound*, ch. 1.

number of potential speech acts might fall under the umbrella of this category – for instance, ‘speech attacks’,⁵⁸ ‘verbal aggression’.⁵⁹ Others define hate speech more narrowly in terms of specific or particular speech acts or disjunctive sets of speech acts – for example, offending or advocating hatred,⁶⁰ insulting or defaming,⁶¹ disparaging,⁶² degrading,⁶³ ambushing, terrorising, wounding, humiliating or degrading,⁶⁴ harming or inciting discrimination or violence,⁶⁵ or inciting persecution.⁶⁶

As I have said, the main aim of this article is to say something on behalf of a broader approach that goes beyond just the legal concept *hate speech*. In order to motivate this, let me briefly note some limitations of the foregoing approach and its results. Aside from the internal inconsistencies and contradictions in the above characterisations, I find them at the same time too narrow (under-inclusive) and too broad (over-inclusive). Because many of the foregoing characterisations are anchored in a particular body of law and legal doctrine, such as an extreme free speech doctrine embraced by many American legal scholars and various United States Supreme Court Justices, or a slightly less extreme free speech doctrine embraced by legislatures and courts in some European countries who are to some extent influenced by these ways of thinking, they are also tethered to the narrow purposes and perspectives of these bodies of law and legal doctrines. An acute awareness of First Amendment protections of freedom of expression and a general presumption against speech prohibitions hangs like a shadow over these characterisations. Consequently, some of the characterisations tend toward highly restrictive understandings of hate speech. The rationale behind narrowing the boundaries of the concept is this: if the legal concept *hate speech* is

⁵⁸ See, e.g., Smolla, *Free Speech in an Open Society*, p. 152; Gould, *Speak No Evil*, p. 14. Cf. Anthony Lewis, *Freedom for the Thought that We Hate* (New York, NY: Basic Books, 2007), p. 157.

⁵⁹ Terry A. Kinney, ‘Hate Speech and Ethnophaulisms’, *The International Encyclopedia of Communication* (Blackwell Reference Online, 2008). Available at http://www.blackwellreference.com/subscriber/uid=/tocnode?id=g9781405131995_chunk_g978140513199513_ss4-1.

⁶⁰ Coliver, ‘Hate Speech Laws: Do They Work?’, p. 363.

⁶¹ Sumner, ‘Hate Crimes, Literature and Speech’, p. 142; Fraleigh and Tuman, *Freedom of Expression in the Marketplace of Ideas*, p. 139.

⁶² Nockleby, ‘Hate Speech’, p. 1277.

⁶³ Kevin W. Saunders, *Degradation: What the History of Obscenity Tells Us About Hate Speech* (New York, NY: New York University Press, 2011), p. 132.

⁶⁴ Lawrence et al., ‘Introduction’, p. 1.

⁶⁵ Schauer, ‘Uncoupling Free Speech’, p. 1349.

⁶⁶ Tesis, *Destructive Messages*, p. 211, n. 1.

to be useful, it should only extend to forms of speech or expression that could conceivably be regulated under a constitutional regime committed to freedom of expression. But this leads to an unnecessarily restrictive account of the ordinary concept *hate speech*, in my view. For example, some writers appear to suggest that something is hate speech if and only if it is incitement to hatred.⁶⁷ Others have even toyed with the idea of restricting what may count as hate speech to speech that is addressed to a captive audience.⁶⁸ These suggestions may be understandable given a particular vision of the function of the legal concept *hate speech*, but they are nevertheless both inaccurate and unhelpful as applied to the ordinary concept.

The above characterisations are also too broad (over-inclusive) in the way they handle the connection between hate speech and hate or hatred. They embrace the myth of hate: the myth that in order for speech to qualify as hate speech, and to be appropriately called 'hate speech' by competent users of the term, the speech must be connected in some relevant way with emotions, feelings, or attitudes of hate or hatred, meaning intense or extreme dislike, aversion, loathing, antipathy, enmity or hostility toward or against something or someone, even an impulse to banish or obliterate that thing or person.⁶⁹ According to Parekh, for example, the word 'hate' in 'hate speech' 'implies hostility, rejection, a wish to harm or destroy, a desire to get the target group out of one's way, a silent or vocal and a passive or active declaration of war against it'.⁷⁰ And this account of the hate in hate speech certainly echoes seminal accounts of *simple hatred* in the philosophy of the emotions.⁷¹ Yet those who embrace the myth of hate do not allow conceptual space for the possibility that emotions, feelings, or attitudes of hate or hatred are merely contingently associated with, and non-essential to, hate speech as

⁶⁷ See, e.g., Cohen-Almagor, *The Scope of Tolerance*, p. 11; Cohen-Almagor, 'Holocaust Denial is a Form of Hate Speech', p. 35; Waldron, 'Dignity and Defamation', p. 1601, n. 14.

⁶⁸ See, e.g., Brink, 'Millian Principles, Freedom of Expression, and Hate Speech', p. 135.

⁶⁹ See, e.g., Glasser, 'Introduction', p. 1; Garner, *Black's Law Dictionary*, pp. 1407–1408; Strossen, 'Incitement to Hatred', p. 244; Corlett and Francescotti, 'Foundations of a Theory of Hate Speech', p. 1083; Fischer, 'First Amendment Morality Versus Civility Morality', pp. 161–162; Parekh, 'Hate Speech', p. 214; Post, 'Hate Speech', pp. 123–125; Weinstein and Hare, 'General Introduction', p. 4; Gagliardone et al., *Mapping and Analysing Hate Speech Online*, p. 9.

⁷⁰ Parekh, 'Hate Speech', p. 214.

⁷¹ See, e.g., Jean Hampton, 'Forgiveness, Resentment and Hatred', in J. Murphy and J. G. Hampton (eds.) *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988), pp. 60–61.

properly understood. The truth is, however, that even though a good deal of hate speech is connected with emotions, feelings, or attitudes of hate or hatred, this is neither inevitable nor necessarily true of all instances of hate speech. For example, some instances of hate speech have nothing to do with hate or hatred, strictly speaking, but everything to do with emotions, feelings, or attitudes of contempt, disdain, scorn, condescension, or dismissiveness, which are matters of regarding something or someone as unworthy of one's consideration or respect and of having an impulse or reason therefore to withdraw from, avoid or shun that thing or person.⁷² In the cases I have in mind the hate speaker does not hate the objects of his speech precisely because he holds them in contempt and, therefore, in his mind, they are beneath a level of consideration or respect that would be required to feel something or have regard for them, be that hate or anything else.⁷³ At least some speech, then, is plausibly counted as hate speech because it expresses or conveys the speaker's feelings or emotions or attitudes of contempt for the objects of the speech as distinct from feelings or emotions or attitudes of hate or hatred.⁷⁴

Note, I do not mean to suggest that hate speech is always connected with hate or contempt or both, and by 'both' I mean something like despisement (hating people because one has contempt for them). If that were true, then we could simply treat the word 'hate' in 'hate speech' as referring to *hate or contempt or despisement* as a disjunctive mental quality that we attribute to the motivations of all hate speakers, and/or something that all hate speech conveys or symbolises, and/or something that all hate speech is designed to produce in the audience (people the hate speaker is seeking to bring on side).⁷⁵ But, in fact, hate speech can be associated

⁷² See, e.g., Michelle Mason, 'Contempt as a Moral Attitude', *Ethics* 113 (2003): 234–272; Kate Abramson, 'A Sentimentalist's Defense of Contempt, Shame, and Disdain', in P. Goldie (ed.) *The Oxford Handbook of Philosophy of Emotion* (Oxford: Oxford University Press, 2009); Macalester Bell, *Hard Feelings: The Moral Psychology of Contempt* (Oxford: Oxford University Press, 2013).

⁷³ I do not claim, however, that hatred and contempt are in themselves mutually exclusive. Cf. Arthur Schopenhauer, *Essays and Aphorisms*, trans. R. J. Hollingdale (London: Penguin Books, 1970), p. 170; Friedrich Nietzsche, 'Epigrams and Interludes', in *Beyond Good and Evil: Prelude to a Philosophy of the Future*, trans. W. Kaufmann (New York: Vintage Books, 1966), p. 92.

⁷⁴ See also Robert Simpson, 'Dignity, Harm, and Hate Speech', *Law and Philosophy* 32: 701–728, pp. 701, n. 2, 702; Robin Jeshion, 'Slurs, Dehumanization, and the Expression of Contempt', in D. Sosa (ed.) *Bad Words* (Oxford: Oxford University Press, forthcoming).

⁷⁵ See, e.g., Raphael Cohen-Almagor, 'Hate and Racist Speech in the United States – A Critique', *Philosophy and Public Issues* 6 (2016): 77–123, pp. 103–104.

with a wide range of different speaker motivations not exhausted by, and potentially not even including, *hate or contempt or despisement*, everything from fear or envy to the need to fit in or a simple desire for attention from others. It can also express or articulate a wide range of different speaker feelings or emotions or attitudes other than *hate or contempt or despisement*, such as anxiety, disgust or even a sense of alienation. And it can be intended to cause or bring about a wide range of different emotional affects or cognitive states in the audience even in the absence of *hate or contempt or despisement*, including, for instance, shock, excitement, pleasure, or even a sense of camaraderie.

By the same token, however, I am not suggesting that the term 'hate speech' is a misnomer in the strict or strong sense. Some hate speech clearly is associated with hate or hatred. Nevertheless, what I am saying is that the term 'hate speech' is highly misleading if the implication is that hate speech is always associated with hate or hatred, and is still somewhat misleading even if the suggestion is that hate speech is normally or typically or in-all-but-outlier-cases associated with hate or hatred.

Now I do think it would be inaccurate, and therefore unfair, to claim that all writers on hate speech fall into the trap of blindly accepting the myth of hate. It seems to me that Catharine MacKinnon, Timothy C. Shiell, Jeremy Waldron, and Martha C. Nussbaum come much closer to avoiding the myth of hate than most writers.⁷⁶ Yet it also strikes me that philosophers should do more than merely beware the myth of hate, they should try to directly challenge it and if possible explode it; and that is precisely what I shall attempt in the next two sections.

III. THE ORDINARY CONCEPT *HATE SPEECH* AS A COMPOSITIONAL CONCEPT?

Let us begin with the idea that the ordinary concept *hate speech* might be a complex or compositional concept. To say that a concept

⁷⁶ See Catharine MacKinnon, 'Pornography as Defamation and Discrimination', *Boston University Law Review* 71 (1991): 793–808, p. 808; Catharine MacKinnon, 'Forward', in I. Maitra and M. McGowan (eds.) *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University Press, 2012), pp. xvi–xvii; Shiell, *Campus Hate Speech on Trial*, pp. 161–162; Waldron, 'Dignity and Defamation', pp. 1600–1601; and Martha C. Nussbaum, 'Haterz Gonna Hate? There Are Limits to What the Law Can do to Police Cyberabuse', *The Nation*, November 5, 2014.

is complex or compositional is to say that it is made from or composed of other, simpler concepts. These simpler concepts are more basic or more ordinary concepts in the sense that they include larger classes of things. Complex or compositional concepts are said to be amenable to a method of *decompositional conceptual analysis*: a process of breaking down the complex concept into its component parts. The standard example is the concept *bachelor* which is said to be made from or composed out of three more basic or more general concepts: (1) *adult human*, (2) *unmarried*, and (3) *male*. Decompositional conceptual analysis is often used to produce precise definitions, the specification of a set of necessary and jointly sufficient conditions for the application of concepts or for the appropriate use of terms used to refer to concepts. That something falls under each of the more basic or more general concepts that compose the compositional concept is necessary and jointly sufficient for that thing to fall under the compositional concept. Thus, it might be said that something falls under the concept *bachelor* only if (1) it is an adult human, (2) it is unmarried, and (3) it is male. And that if something is (1) an adult human, (2) unmarried, and (3) male, then it is a bachelor.

To apply this method to the ordinary concept *hate speech* is to attempt to break down this concept into its component or constituent parts, its more basic, simpler concepts and to use these concepts to produce a precise definition. If the ordinary concept *hate speech* is a compositional concept, how are its component parts to be identified? One way is to focus on the words that make up the term 'hate speech'. In particular, one could work on the assumption that this term is not simply a compound noun in its grammatical or syntactical form⁷⁷ but also a compound noun in its meaning or semantic content. In short, one could treat the meaning of the term 'hate speech' as being a function of the meanings of its constituent parts. Following on from this, one might suppose that in order to divine the meaning of the term 'hate speech' one simply understands the meanings of the two words 'hate' and 'speech' and links those

⁷⁷ One interesting grammatical or syntactic feature of compound nouns is that it is possible to stack two or more nouns together or construct compound nouns recursively. So one can add together 'fiction' and 'science' to make 'science fiction' and then combine 'writer' and 'science fiction' to make 'science fiction writer'. Similarly, one can add together 'speech' and 'hate' to make 'hate speech' and then combine 'law' and 'hate speech' to make 'hate speech law'.

meanings to each other in some appropriate way.⁷⁸ For example, one might suppose that the term ‘hate speech’ is a hyponym of the word ‘speech’. So just as the term ‘olive oil’ denotes a subcategory of oil, the term ‘hate speech’ denotes a subcategory of speech. What is more, it might be thought that the word ‘hate’ in ‘hate speech’ functions semantically to tell us something about the essential nature of the subcategory of speech in question, namely, that it involves or is intimately connected with emotions, feelings, or attitudes of hate or hatred. Indeed, the fact that the compound noun ‘political speech’ is semantically compositional (it denotes a subcategory of speech that is political in nature) may lead us to assume that the compound noun ‘hate speech’ adheres to the same principle.

So far I have been assuming that the ordinary concept *hate speech* is composed of the two more basic concepts *hate* and *speech*. However, on further reflection we might think that the concept *hate speech* is even more complex than the words ‘hate’ and ‘speech’ imply. Perhaps the concept *hate speech* is also composed of other concepts. One possibility is the concept *group or class of persons*. Most people would instinctively recognise as hate speech, I think, the slogan of the nineteenth century nativist, Denis Kearney: ‘To an American, death is preferable to life on a par with the Chinaman’.⁷⁹ Part of the reason we would recognise this as hate speech is that it is about or against Chinese people. Similarly, few people nowadays would fail to understand that between the two insults, ‘Fuck you!’ (one white person to another white person in anger) and ‘Fuck you, nigger!’ (one white person to a black person in anger), the latter is more likely to have the label ‘hate speech’ attached to it. Of course, a racial insult, such as this, can be, and often is, targeted at or directly addressed to only one individual. But where the insult takes the form of a pejorative reference to the individual’s race, the individual may be no more than a representative of an entire group or class of persons in the mind of the hate speaker. If the addressee is interchangeable, in that sense, then the hate speech is really about or against the entire group or class of persons.

⁷⁸ See, e.g., David Boromisza-Habashi, *Speaking Hatefully: Culture, Communication, and Political Action in Hungary* (University Park, PA: Penn State University Press, 2013), pp. 35–36; Gagliardone et al., *Mapping and Analysing Hate Speech Online*, p. 9.

⁷⁹ Cited in Elmer Sandmeyer, *The Anti-Chinese Movement in California* (Chicago, IL: University of Chicago Press, 1973), p. 65.

Things might not be so straightforward, however. When critical race theorists started to employ the term ‘hate speech’ they did so in relation to speech about or against not just any type of group or class of persons. They classified racist insults like, ‘Fuck you, nigger!’ (a white person to a black person in anger), as hate speech precisely because of the oppressive social relationships embodied and facilitated by this sort of speech.⁸⁰ Thus, Matsuda classifies as ‘hate speech’ a case in which a group of white workers place a noose in the work area of an African-American worker,⁸¹ whereas she interprets ‘an angry, hateful poem by a person from a historically subjugated group as a victim’s struggle for self-identity in response to racism’.⁸² Following this lead, some writers have analysed the concept *hate speech* as being partly composed of the concept *historically oppressed or victimised groups or classes of persons*.⁸³ According to this analysis, if, taking into account other tell-tale signs (to be discussed in ‘Family Resemblances’), certain words or symbols are about or against members of a historically oppressed or victimised group, then this can lend credence to words or symbols being hate speech. Then again, more recent usage seems to permit a competent user to describe as ‘hate speech’ instances in which members of an oppressed or victimised group lashes out verbally against members of an oppressive or victimising group – ‘Fuck you, white devil!’⁸⁴ This newer accepted usage may partly reflect a white backlash against the original use of the term ‘hate speech’ (to condemn what whites had to say about other racial or ethnic groups), but it is also likely to reflect the broader work that society as a whole now expects the term ‘hate speech’ to do – for example, to identify or flag up forms of speech that society as whole fears have the potential to be very socially divisive or destructive of social cohesion. I shall expand on some of the broader functions of the term ‘hate speech’ in ‘Family Resemblances’. Therefore, maybe the more basic concept here is *groups or classes of persons identified by protected characteristics*.

⁸⁰ See, e.g., Matsuda, Matsuda, ‘Public Response to Racist Speech’, p. 2357.

⁸¹ *Ibid.*, p. 2327.

⁸² *Ibid.*, pp. 2361–2362.

⁸³ See, e.g., Cohen-Almagor, *The Scope of Tolerance*, p. 153; David Boromisza-Habashi, ‘Hate speech’, *The International Encyclopedia of Language and Social Interaction* (Boston, MA: John Wiley and Sons, 2015), p. 715.

⁸⁴ See, e.g., Greenawalt, *Fighting Words*, p. 164, n. 39.

But what are protected characteristics? Many writers seem to define this concept by simply enumerating a list of characteristics, statuses or identities that they take to be examples, such as race, ethnicity, religion, nationality, gender or gender identity, sexual orientation, and disability.⁸⁵ But this cannot count as an adequate definition: first, because the lists proffered are not, and are not intended to be, exhaustive, merely illustrative or indicative; second, because it is rarely explained what feature or essential nature is shared by all of the characteristics listed. It might be tempting at this point to define the term ‘protected characteristics’ in a legalistic way, as characteristics that are legally protected. But this seems to create a status quo bias in favour of whatever characteristics are protected by current hate speech laws. It renders problematic statements of the following form: ‘Hate speech against group x isn’t currently, but ought to be, banned because x should be a protected characteristic.’ This statement appears meaningful, but if ‘hate speech’ means speech that picks out protected characteristic, and the statement is deliberately referring to a characteristic that is not currently protected, then strictly speaking the statement is meaningless.

An alternative is to define the term ‘protected characteristics’ as characteristics that ought to be legally protected. But this definition relies on a deeper account of the real or underlying function or purpose of hate speech law, that is, a justification of why we should have hate speech law in the first place. With such an account in hand we might then be in a position to explain why some things should be on the list of protected characteristics and why other things should not. A right-wing politician might choose to express in public that he sincerely hates communists, for example, but we might think it inappropriate to label this ‘hate speech’ given the real or underlying function of hate speech law. Identifying or flagging up speech that attacks people on grounds of their political or ideological beliefs is arguably *not* one of the important jobs or roles that we have hitherto assigned to or carved out for the legalised concept *hate speech*.

⁸⁵ See, e.g., Brison, ‘The Autonomy Defense of Free Speech’, p. 313; Delgado and Stefancic, *Understanding Words That Wound*, p. 11. For a discussion of the numerous characteristics that governments could potentially deem ‘protected’ for the purposes of hate speech law, see Alexander Brown, ‘The ‘Who?’ Question in the Hate Speech Debate: ‘Part 1’: Consistency, Practical, and Formal Approaches’, *Canadian Journal of Law & Jurisprudence* 29 (2016): 275–320; and Alexander Brown, ‘The ‘Who?’ Question in the Hate Speech Debate: Part 2: Functional and Democratic Approaches’, *Canadian Journal of Law & Jurisprudence* 30 (2017): 23–55.

So let us assume for the sake of argument that the concept *hate speech* is composed of the following three more basic, simpler concepts: (1) *speech or other expressive conduct*, (2) *groups or classes of persons identified by protected characteristics*, and (3) *emotions, feelings, or attitudes of hate or hatred*. This permits an analysis of the ordinary concept *hate speech* in terms of three necessary conditions for something being an instance of the concept. Under this analysis, something is hate speech only if it (1) is speech or expressive conduct, (2) concerns any members of groups or classes of persons identified by protected characteristics, and (3) involves or is intimately connected with emotions, feelings, or attitudes of hate or hatred. What is more, this set of necessary conditions itself constitutes a sufficient condition for something counting as hate speech. Thus, if something qualifies as (1), (2), and (3), then it is hate speech. Let us also assume for the sake of argument that this analysis can shed light on the meaning of the compound noun 'hate speech', whether it is understood syntactically as a noun–noun or as an adjective–noun, compound noun.

In the remainder of this section, I want to focus on the third part of this conceptualisation: the basic concept *emotions, feelings, or attitudes of hate or hatred* as a building block of the putatively complex concept *hate speech*. As mentioned above, an assumption that is explicit or implicit in the work of quite a number of writers on hate speech is that semantically the word 'hate' carries its ordinary or literal meaning: it denotes the presence of emotions, feelings, or attitudes of hate or hatred.⁸⁶ The myth of hate certainly takes us some of the way to understanding condition (3) above, in the sense that it reminds us of the ordinary or literal meaning of 'hate'. But it still leaves unanalysed the precise sense in which speech *involves* or is *intimately connected* with emotions, feelings, or attitudes of hate or hatred. In principle, I believe that this involvement or connection could take at least four possible forms.⁸⁷ However, I shall try to show for each form that the connection between hate speech and emotions, feelings, or attitudes of hate or hatred is contingent.

⁸⁶ See, e.g., Glasser, 'Introduction', p. 1; Garner, *Black's Law Dictionary*, pp. 1407–1408; Strossen, 'Incitement to Hatred', p. 244; Corlett and Franciscotti, 'Foundations of a Theory of Hate Speech', p. 1083; Fischer, 'First Amendment Morality Versus Civility Morality', pp. 161–162; Parekh, 'Hate Speech', p. 214; Post, 'Hate Speech', pp. 123–125; Weinstein and Hare, 'General Introduction', p. 4; Gagliardone et al., *Mapping and Analysing Hate Speech Online*, p. 9.

⁸⁷ Boromisza-Habashi proffers three. See his *Speaking Hatefully*, pp. 35–36.

A. *The Reaction of Hearers*

One possible connection between speech and hate concerns the reaction of hearers. Here the presence of hate speech is signposted by hatred as a reactive emotion, by the way that hearers, including not only the objects or targets of hate speech but also its audiences, feel about hate speech when they hear or learn about it. On this analysis, something is hate speech only if: (3a) it arouses reflexive hatred amongst a significant section of society directed toward the speech and perhaps also the speaker. In other words, the putative connection between speech and hate consists in the fact that a significant section of society hates certain forms of speech that are about or against groups or classes of persons identified by protected characteristics.

This connection echoes one possible analysis of the related concept *hateful speech*. There are two main senses of the word 'hateful'. In one sense, the word 'hateful' signals something about how the hearer regards the speech in question, as speech that is detestable, repugnant or despicable, so much so that it arouses emotions, feelings, or attitudes of hate or hatred in the hearer toward the speech and perhaps also the speaker, or deserves to arouse such hatred. In a second sense the word 'hateful' connotes something about the content of the speech itself, as speech that says something derogatory or deeply insulting about members of certain groups. I shall return to this second sense in 'Family Resemblances'. Now sometimes when people call the word 'nigger' a 'hateful expression', or describe acts of burning crosses outside the homes of African American families or marching in Jewish neighbourhoods wearing replica Nazi uniforms and Swastika symbols as 'hateful acts', they have in mind the first sense of 'hateful'. They mean that such words or expressive acts are so detestable that they either do or should arouse emotions, feelings, or attitudes of hatred in any fair-minded person.⁸⁸

As a conceptual analysis of the term 'hate speech', however, the current connection has few serious advocates, Nadine Strossen being

⁸⁸ Some writers use the words 'hate speech' and 'hateful speech' interchangeably, as though they are mere synonyms, and, what is more, fail to distinguish between different senses of 'hateful'. See, e.g., Richard Delgado and Jean Stefancic, *Must We Defend Nazis? Hate Speech, Pornography, and The First Amendment* (New York, NY: New York University Press, 1997), p. 128; Butler, *Excitable Speech*, p. 102; Ian Cram, 'Hate Speech and Disabled People: Some Comparative Constitutional Thoughts', in A. Lawson and C. Gooding (eds.) *Disability Rights in Europe: From Theory to Practice* (Oxford: Hart Publishing, 2005), pp. 65–67; Heinze, *Hate Speech and Democratic Citizenship*, p. 19.

an exception.⁸⁹ This is for good reason. For one thing, there may be things that (1) are forms of speech, (2) are about or against members of groups or classes of persons identified by protected characteristics and (3a) arouse reflexive hatred amongst some significant section of society, but which nevertheless do *not* seem to fall under the ordinary concept *hate speech*. Suppose a controversial politician in a majority white racist society makes a speech arguing for the equal rights of blacks. Even if this speech satisfies conditions (1), (2) and (3a), it scarcely seems plausible to call it 'hate speech'. On the other hand, there could be things that fail to (3a) arouse reflexive hatred amongst some significant section of society, but which nevertheless *do* seem to fall under the ordinary concept *hate speech*. Consider another politician in the majority white racist society who makes an extremely subtle, veiled speech about inherent differences between racial groups, one that fails to arouse emotions, feelings, or attitudes of hate or hatred against the speech. We might yet call this 'hate speech' even if people do not hate it. The point is that hearer reactions, including hate or hatred, are not always reliable guides to how we should use the term 'hate speech'.

What is more, in many countries or socio-legal communities hate speech is treated as regulatable speech. I shall say more about this in the 'Family Resemblances'. But if we define hate speech as speech that audiences hate because of what it says about members of certain groups, then this would seem to undermine the legitimacy of the idea that hate speech is regulatable speech. This is because we tend to think the mere fact that an audience hates a form of speech is not in itself a legitimate reason to ban it.⁹⁰ Now it might be objected at this point that to reject the present way of understanding the ordinary concept *hate speech* on the proposed grounds would be to once again allow the legal concept *hate speech* to cast a shadow over the analysis. In other words, because the legal concept *hate speech* eschews any suggestion that hate speech is simply *speech we hate* because of legal sensibilities over what is regulatable speech, this deters us from understanding the ordinary concept *hate speech* in those terms. But the main thrust of my arguments in these articles is not toward the conclusion that the legal and ordinary concepts must

⁸⁹ Strossen, 'Incitement to Hatred', p. 244.

⁹⁰ Cf. Richard Delgado and David Yun, "'The Speech We Hate': First Amendment Totalism, the ACLU, and the Principle of Dialogic Politics', *Arizona State Law Journal* 27 (1995): 1281–1300.

always be kept apart and that appealing to any analyses of the former can only detract from or thwart good analyses of the latter. Rather, my point is that *some* aspects of the analysis of the legal concept *hate speech* may be unhelpful to understanding the ordinary concept *hate speech*. For example, to refuse to admit as examples of hate speech anything that would be protected from regulation under a relatively extreme free speech doctrine embraced by many American legal scholars and United States Supreme Court justices would be unhelpful, especially when the ordinary concept *hate speech* extends well beyond that doctrine. By contrast, to reject as a conception of hate speech the idea that hate speech is simply *speech we hate* because it would be inconsistent with virtually any free speech doctrine embraced by legal systems around the world would be appropriate. Clearly both the legal and the ordinary concepts *hate speech* will draw on, and be infused by, a shared body of general ideas about what is, and what is not, regulatable speech that can be found in the wider socio-legal context or cultural milieu.

B. What Hate Speakers are Expressing

A second possible explanation for the connection between speech and hate focuses on what hate speakers are expressing. On this analysis, something is hate speech only if: (3b) it expresses the emotions, feelings, or attitudes of hate or hatred of the speaker for any members of groups or classes of persons identified by protected characteristics.⁹¹ Another word for this sort of speech might be 'hate-filled speech' or 'hate-laden speech'. It is speech that carries forth, articulates, symbolises or represents in some way the speaker's emotions, feelings, or attitudes of hate or hatred toward the subject of the speech. An example of hate-laden speech in its explicit form might include an utterance that began, 'I hate...', as in 'I hate those damn Jews!'. However, that a speaker chooses to use or not use the words 'I hate...' is not decisive. Most hate speakers are only too aware of hate speech laws or regulations and sometimes prefix their

⁹¹ See note 44 above.

utterances with 'I honestly don't hate x, but...' in an attempt (probably misguided) to insulate themselves from prosecution.⁹² Therefore, it is important to recognise that hate-laden speech can take implicit forms, such as when people express their hatred of certain groups in subtle, careful, tendentious, disguised or dissembling ways (without ever explicitly saying, 'I hate...').

Once again, however, this analysis is flawed. For one thing, not everything that seems to fall under the concept *hate speech* is an expression, explicit or implicit, of emotions, feelings, or attitudes of hate or hatred toward members of groups or classes of persons identified by protected characteristics. Consider four examples. First, suppose a fundamentalist or evangelical Christian directs the following words at people on a public street as they enter an LGBT social club, 'You homosexuals and lesbians are sinners in the eyes of God, you disobey the teachings of the Bible, and for this reason you will go to hell if you do not repent.' It is surely conceivable that this utterance could function not as an articulation or expression of emotions, feelings, or attitudes of hate or hatred but rather as an expression of deeply held religious beliefs or as some sort of religious exercise or rite of passage or even as an outpouring of feelings or sentiments of disgust or repulsion that have been learned from parents or community leaders. If (3b) is a necessary condition for hate speech, then this example is not hate speech. But either as a form of speech that carries the prejudiced (because contemptuous) message that homosexuals are morally inferior beings or as a speech act that amounts to ranking, degrading, harassing or persecuting

⁹² In March 2004, for example, the BNP's Tony Collett made a speech at a public house in Keighley in which he allegedly stated 'I honestly don't hate asylum seekers – these people are cockroaches and they're doing what cockroaches do because cockroaches can't help what they do, they just do it, like cats meow and dogs bark'. Guardian Staff and Agencies, 'Jury Hears of BNP's 'Multiracial Hell' Speech', *The Guardian*, November 3, 2006. If this was a strategy to avoid prosecution it reveals a certain ignorance of the law. Putting to one side the question of whether asylum seekers could be counted as a racial group, in the UK it is not against the law to express hatred for racial groups *per se*. Rather, it is against the law to stir up hatred against racial groups.

homosexuals, it might qualify as hate speech nevertheless.⁹³ Second, suppose a scientist publishes his controversial findings concerning the relative intelligence of African Americans including the statement, 'African Americans tend to have lower IQs than white Americans.'⁹⁴ Again, it is quite possible that his statement is not an expression of any emotions, feelings, or attitudes of hate or hatred toward African Americans, conscious or unconscious. Instead, it could be merely the expression of a sincerely held scientific belief about the effect of both genes and environment on interracial differences in IQ, and about how IQ is a good indicator of interracial differences in such things as educational attainment, income, and reproductive and marriage choices. Because of this there may be an impulse not to count it as hate speech.⁹⁵ Yet it is also perfectly intelligible for people, not only African Americans, to think that this is hate speech by virtue of the fact that in a very public way it

⁹³ Of course, it is a further question whether or not this type of hate speech should be regulated all things considered given the countervailing right to freedom of religion. See, e.g., Albert Weale, 'Freedom of Speech vs. Freedom of Religion', in *Free Speech: Report of a Seminar. Discussion Papers 2* (London: Commission for Racial Equality, 1990); Robert Wintemute, 'Religion vs. Sexual Orientation: A Clash of Human Rights?', *Journal of Law and Equality* 1 (2002): 125–154; Ian Leigh, 'Hatred, Sexual Orientation, Free Speech and Religious Liberty', *Ecclesiastical Law Journal* 10 (2008): 337–344; Ian Leigh, 'Homophobic Speech, Equality Denial, and Religious Expression', in I. Hare and J. Weinstein (eds.) (2009) *Extreme Speech and Democracy* (Oxford: Oxford University Press, 2009); Carolyn Evans, 'Religious Speech that Undermines Equality', in I. Hare and J. Weinstein (eds.) (2009) *Extreme Speech and Democracy* (Oxford: Oxford University Press, 2009); Ashby Jones, 'Is Hateful Speech at a Military Funeral Protected by the First Amendment?', *Wall Street Journal*, August 31, 2010; Jeffrey Shulman, 'When is Religious Speech Outrageous? Snyder v. Phelps and the Limitations on Religious Advocacy', *Penn Statim* 114 (2010): 13–18; Adelle Banks, 'Is Phelps a role model on free speech issues?', *Christian Century* 128 (2011): 18; Alon Harel, 'Hate Speech and Comprehensive Forms of Life', in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012); Brown, *Hate Speech Law*, pp. 74–75, 93–94. In Canada, for example, the part of the criminal code that bans wilful promotion of hatred also contains exemptions or permissible defences against prosecution 'if, in good faith, the person expressed or attempted to establish by an argument an opinion on a religious subject or an opinion based on a belief in a religious text'. See s. 319(3)(b) of the Criminal Code (as amended by An Act to Amend the Criminal Code (Hate Propaganda) of 2004). Similarly, some states in Australia have on the books legislation banning incitement to hatred which sets out exemptions for speech that has a religious purpose or is motivated by sincerely held religious beliefs. See ss. 11(b)(i) and 11(2) of the Racial and Religious Tolerance Act 2001 (the State of Victoria) and see s. 80G(1)(b)(i) and 80G(1)(b)(i) (as amended by s. 6 of Law No. 80 of 2004) (the State of Western Australia). In the UK, s. 29JA of the Public Order Act 1986 (as amended by s. 74 of the Criminal Justice and Immigration Act in 2008 and by Schedule 7 of the Marriage (Same Sex Couples) Act 2013) might be interpreted as implicitly giving special dispensation *inter alia* to religious hate speech concerning sexual conduct or practices and same-sex marriage.

⁹⁴ This need not be a hypothetical example. The authors of *The Bell Curve: Intelligence and Class Structure in American Life*, for example, assert that the average IQ of African Americans is 85, Latinos 89, Whites 103, Asians 106, and Jews 113. Richard J. Herrnstein and Charles Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (New York, NY: Free Press, 1994).

⁹⁵ See, e.g., Mari Matsuda, 'Commentary—Language as Violence v. Freedom of Expression: Canadian and American Perspectives on Group Defamation', *Buffalo Law Review* 37 (1988–1989): 359–364, p. 364.

rehearses and supports, rather than challenges and confronts, a particular negative stereotype or social stigma about African Americans as unintelligent or less intelligent than whites.⁹⁶ Third, it is possible that some disablists could use the insults ‘retard’, ‘cripple’, and ‘freak’ in order to express their contempt or disdain for people with physical or mental disabilities, to express the fact that they look down on such people, believing them to be worthless or beneath consideration, without at the same time articulating, symbolising or representing emotions, feelings, or attitudes of hate or hatred toward such people.⁹⁷ They may even think that hate or hatred would be wasted on ‘pitiable creatures’ whom they believe are less than fully human. But even if such uses of ‘retard’, ‘cripple’, and ‘freak’ express emotions, feelings, or attitudes of contempt as opposed to hate, it does seem perfectly meaningful to call these uses ‘hate speech’. Fourth, consider a group of married men talking together about their wives and women in general. Together they espouse a range of derogatory ideas and negative stereotypes about women (e.g., ‘a woman’s place is in the home’, ‘the man is in charge’, ‘women are no good at practical things around the house’), and at the same time they neither question these ideas nor offset them with more positive stereotypes.⁹⁸ But just suppose for the sake of argument that the men do not in fact harbour any emotions, feelings, or attitudes of hate or hatred toward their wives, meaning they do not possess intense or extreme dislike of their wives and do not have any attitudes of enmity or hostility toward their wives and so do not regard them as people to be banished or destroyed. Nonetheless, we might instinctively think that this sort of sexist discourse can be appropriately classified as hate speech, perhaps because it expresses a certain type of disdain or condescension toward women or maybe because it

⁹⁶ See, e.g., Bob Herbert, ‘In America; Throwing a Curve’, *The New York Times*, October 26, 1994. Once again, it is a further question whether or not this particular type of hate speech should be regulated all things considered. It might be argued that hate speech laws should include exemptions or permissible defences relating to scientific speech for fear of impeding scientific truth discovery, not because scientific speech can never be hate speech but because when push comes to shove not all hate speech should be banned. In some countries, such as Australia and South Africa, there are certain laws banning hate speech that include exemptions or permissible defences for scientific speech. See s. 18D(b) of the Racial Discrimination Act 1975 (Australia), s. 20C(2)(c) of the Anti-Discrimination Act 1977 (New South Wales), s. 66(2)(c) of the Discrimination Act 1991 (Australian Capital Territory), and s. 12 of the Promotion of Equality and Prevention of Unfair Discrimination Act 2000 (South Africa).

⁹⁷ Cf. Robin Jeshion, ‘Expressivism and the Offensiveness of Slurs’, *Philosophical Perspectives* 27 (2013): 231–259, p. 241.

⁹⁸ Note, I have given an example of some men talking about women but the example is not intended to depict or stereotype all men.

serves to maintain forms of gender subordination (from which men benefit).⁹⁹

Of course, a defender of the view that hate speech necessarily has something to do with expressing the speaker's feelings or emotions or attitudes of hate or hatred, could simply deny that the above are genuine examples of hate speech. They might insist that as soon as we move beyond the relatively narrow and precise legal *concept* and begin to carve out the boundaries of an ordinary concept *hate speech*, we are in uncharted waters, so we must expect people to use the term in inconsistent or even chaotic ways, or perhaps to use the term in ways that merely serve political or ideological ends ('I call this hate speech because this sort of speech offends me or contradicts what I believe'). And so (according to this response), the person who uses the term 'hate speech' to describe the above examples, even though the speech does not express hate or hatred, is simply confused or an incompetent user of the term, or perhaps is using the term in a deliberately subversive or provocative way.

However, I think that this dismissal of the counter-examples is too quick. For one thing, if one is relying on the legal concept *hate speech* to provide some sort of guidance for the task of carving out an ordinary concept *hate speech*, then we must also pay attention to the fact that most legal regimes that actually do employ the term 'hate speech' do so in relation to incitement to hatred legislation, wherein the necessary mental element of the relevant offences tends to be intent, as opposed to the expression of emotions, feelings, or attitudes of hate or hatred. Moreover, if the current version of the myth of hate were true, then ordinary users of the term 'hate speech' would tend to look upon the following sort of statement as involving considerable confusion on the part of the speaker, as involving a contradiction in terms. 'What the scientist said was hate speech; now I'm not suggesting that the speech was expressing any feelings or emotions or attitudes of hate or hatred, I strongly suspect it wasn't, but it was hate speech nevertheless because of the way it negatively stereotyped people of colour.' But it seems to me that most ordinary language users would, or could, find this sort of statement quite meaningful and not a contradiction in terms, regardless of whether they happened to agree or disagree with it.

⁹⁹ See, e.g., Donna L. Lillian, 'A Thorn By Any Other Name: Sexist Discourse as Hate Speech', *Discourse Society* 18 (2007): 719–740, pp. 732–733.

Furthermore, arguably carving out the boundaries of a new ordinary concept *hate speech* is partly a matter of trying to say something about what the term 'hate speech' ought to mean if it is to be useful or do any good. But arguably, conceptualising hate speech as the expression of hate or hatred may have the unwelcome effect of pathologising hate speech. To conceptualise hate speech as the act of people who harbour so much hatred or deep-seated animosity toward members of certain groups that they cannot help themselves but express their hatred outwardly through speech, seems to imply that hate speech is abnormal or a symptom of mental pathology. Hate speech becomes the act of someone with obsessional, paranoid, or irrational feelings or emotions or attitudes and uncontrollable impulses to express these things to other people. However, we might intuitively think that a great deal of actual hate speech is undertaken by non-pathological people – people who may not harbour hatred or deep-seated animosity toward others, and who, even if they do harbour such hatred, are perfectly capable of resisting outwardly expressing it through speech.¹⁰⁰

There is another potential drawback. Consider once again the proposition that hate speech is a type of regulatable speech – a proposition that is perhaps a function of the connections between the ordinary and legal concepts *hate speech*. The current way of defining hate speech seems to be incompatible with this proposition. After all, if what makes something hate speech is that it expresses the speaker's emotions, feelings, or attitudes of hate or hatred, then there is potential that any laws designed specifically to prohibit hate speech *so defined* would thereby also prohibit, by design, potentially important forms of self-expression and self-realisation. For example, it might be that one way of coming to know one's own emotions, feelings, or attitudes is through expressing or articulating them in

¹⁰⁰ I take inspiration here from any argument made by Barbara Perry about conceptualisations of the concept of hate crime that emphasise feelings of hatred. Barbara Perry, 'A Crime by Any Other Name: The Semantics of 'Hate'', *Journal of Hate Studies* 4 (2005): 121–137, p. 124.

public.¹⁰¹ Not only that, one might also take the view that expressing hatred in public acts as a safety-valve, a way of letting off emotional steam, so to speak, as an alternative to acting out emotions, feelings, or attitudes of hate or hatred through acts of discrimination or violence.¹⁰² (Some scholars reject certain key premises of the safety-valve argument, of course,¹⁰³ but let us just assume for now the argument is sound.) If so, then potentially any laws designed specifically to prohibit hate speech *so defined* would thereby also prohibit, by design, potentially important forms of letting off emotional steam. All of this might seem to suggest that hate speech is not regulatable speech after all. Therefore, if one is committed to the proposition that hate speech is regulatable speech, then one has reason to refrain from defining hate speech as the expression of hate or hatred.

C. *The Motivation of Hate Speakers*

A third possible connection between speech and hate has to do with the motivation of speakers for saying what they say. On the proposed analysis, something is hate speech only if: (3c) it is motivated by the speaker's hatred for any members of groups or classes of persons identified by protected characteristics.¹⁰⁴ We might call this 'hate-fuelled speech'. This is speech that is motivated by the speaker's hatred of a given group of people, irrespective of whether or not it serves to express or represent that hatred. As in the case of hate-laden speech, hate-fuelled speech is identified by the state of mind or

¹⁰¹ See, e.g., Samuel P. Nelson, *Beyond the First Amendment: The Politics of Free Speech and Pluralism* (Baltimore, MD: The Johns Hopkins University Press, 2005), p. 65. Cf. Brown, *Hate Speech Law*, pp. 120–127.

¹⁰² See, e.g., Calvert Magruder, 'Mental and Emotional Disturbance in the Law of Torts', *Harvard Law Review* 49 (1936): 1033–1067, p. 1053; Thomas Emerson, 'Toward a General Theory of the First Amendment', *Yale Law Journal* 72 (1963): 877–956, pp. 884–885; Marjorie Heins, 'Banning Words: A Comment on "Words That Wound"', *Harvard Civil Rights-Civil Liberties Law Review* 18 (1983): 585–592, p. 590.

¹⁰³ See, e.g., Richard Delgado, 'Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling', *Harvard Civil Rights-Civil Liberties Law Review* 17 (1982): 133–181, p. 140; Richard Delgado and David Yun, 'Pressure Valves and Bloodied Chickens: An Assessment of Four Paternalistic Arguments for Resisting Hate-Speech Regulation', *California Law Review* 82 (1994): 871–892, pp. 878–880; Richard Delgado and Jean Stefancic, 'Ten Arguments Against Hate-Speech Regulation: How Valid?', *Northern Kentucky Law Review* 23 (1996): 475–490, pp. 477–479.

¹⁰⁴ See note 45 above.

the emotional state of the speaker. But whereas the expressive content of hate-laden speech directly represents or articulates the hatred of the speaker, hate-fuelled speech is a function of the motivation (hatred) of the speaker. On this analysis, the term 'hate speech' might be interpreted as an adjective-noun compound noun meaning *speech that is motivated by hatred*.

The task of identifying actual examples of hate-fuelled speech is made difficult by a lack of reliable information about the speaker's state of mind. In the absence of this information, we may be left drawing inferences from what was said, the circumstances in which it was said, and the personal history of the speaker. In other words, we infer that the speaker would simply not have said what he said about members of certain groups unless he possessed, even if unconsciously, emotions, feelings, or attitudes of hate or hatred for them. For example, some people take the view that in the context of the overwhelming weight of evidence that the Holocaust happened, Holocaust denial must be motivated by an implicit hatred of Jews – a hatred that leads or motivates Holocaust deniers to misread, misconstrue, ignore, overlook, twist, and distort the evidence.¹⁰⁵ In other words, some people take the view that but for hatred of Jews, nobody would deny the Holocaust. If this is correct, then Holocaust denial is hate-fuelled speech. Such speech is motivated by hatred even if its expressive content does not directly represent or articulate that hatred.

There are, once again, difficulties with this analysis in addition to problems of identifying actual examples. When something is identified as hate speech by virtue of a particular mental element that is thought to be present in the motivational structure of the speaker, this opens up the possibility of things being conceptualised as hate speech even if they do not portray any of the normal or intuitive signs of being hate speech. Suppose a senior politician or superior judge in a racist society utters the words, 'I hereby grant equal voting rights to blacks.' But suppose his motivation for doing so is actually an extreme hatred of blacks. Suppose he believes that once blacks have the vote they will make such poor decisions that the country will descend quickly into bankruptcy and civil war, causing blacks

¹⁰⁵ This view is articulated on the website of the United States Holocaust Memorial Museum, for example. 'Holocaust Denial and Distortion', Available at www.ushmm.org/confront-antisemitism/holocaust-denial-and-distortion.

tremendous hardship and suffering. So powerful is his hatred that he is prepared to cut off his nose to spite his face. His utterance satisfies conditions (1), (2) and (3c), but could we appropriately call it 'hate speech'?

Conversely, there may be forms of speech that we would intuitively want to say are examples of hate speech even if they are not motivated by hatred. Someone could choose to publicly insult or deride members of the Muslim community where she lives, for example, not out of any emotion or feeling or attitude of hate or hatred but out of a sense of fear or loss or alienation caused by the presence of any 'foreigners' in *her* community. Usage of what we instinctively want to label 'hate speech' might be driven by a range of motives which not merely have nothing to do with hatred *per se* but also have nothing intrinsically to do with the groups in question. Boredom, attention-seeking, pleasure in being controversial, or even something as mundane as economic self-interest – these are just some of the generic motives that may explain the behaviour of hate speakers in some circumstances. Consider the shopkeeper who makes false damaging statements of fact about Jews living in his area not because he hates Jews – in fact he greatly admires Jews – but because he sees Jewish-owned businesses as economic competitors and he wishes to steal their customers with the help of a smear campaign.

In addition to this, there is the thought that hate speech is regulatable speech. The potential disadvantage with defining hate speech as speech motivated by hate or hatred is that any laws that are designed to regulate hate speech so defined could be seen as, in effect, regulating people's emotions, feelings, or attitudes. Yet many people instinctively believe that it is not the job of the state to regulate people's inner or private world. The state should no more legislate emotions and feelings than it should legislate attitudes or thoughts, even if they involve hate or hatred. Indeed, it is partly because people believe a free society is one in which legislators should not have the prerogative to use laws for the purposes of thought control that people support the constitutional right to freedom of conscience. Thus, one danger with analysing the word 'hate' in 'hate speech' in the proposed way is, as Waldron points out, that it gives the clear impression 'that the task of legislation that restricts hate speech is to try to change people's attitudes or control

their thoughts'.¹⁰⁶ Of course, there are advocates of hate speech regulations (not to mention hate crime or penalty-enhancement regulations) that have actively embraced this impression.¹⁰⁷ Then again, there is surely also a case to be made that creating this impression severely weakens the claim that hate speech is a type of regulatable speech. Indeed, because civil libertarians wrongly assume that the concept *hate speech* is really about the attempt to legislate thoughts, they argue that there is a slippery slope that begins with the prohibition of hate speech and ends with the prohibition of 'hate thought', a type of thought-crime in the Orwellian sense.¹⁰⁸ So as well as linguistic reasons not to define 'hate speech' in terms of a motivation of hate or hatred, there may also be sound purpose-oriented reasons not to do so. More on such reasons in 'Family Resemblances'.

D. *The Effects of Hate Speech*

A fourth possible connection between speech and hate highlights the emotions, feelings, or attitudes of hate or hatred that can be incited or stirred up by speech, whether by intention or by likelihood. According to Waldron, for example, 'if 'hatred' is relevant at all, it is relevant (in many of the statutory formulations) as the purpose of the offending speech, not as its motivation'.¹⁰⁹ On this analysis, something is hate speech only if: (3d) it incites hatred toward any members of groups or classes of persons identified by protected characteristics.¹¹⁰ Call this 'hatred-inciting speech' or 'incitement to hatred'. Like the first connection outlined above, incitement to hatred involves the arousal of hatred in the audience or hearer. But whereas in the case of the first connection the hatred aroused rebounds back toward the speech itself or the speaker, in the case of incitement to hatred the hatred is aroused or stirred up in the audience or hearer toward members of some other group or class of persons. To clarify, the supposed connection between speech and

¹⁰⁶ Waldron, 'Dignity and Defamation', p. 1601.

¹⁰⁷ See, e.g., Delgado, 'Words That Wound', pp. 148–149; Diana Tietjens Meyers, 'Rights in Collision: A Non-Punitive, Compensatory Remedy for Abusive Speech', *Law and Philosophy* 14 (1995): 203–243, p. 233.

¹⁰⁸ Joseph Fallon, 'UK-USA: The British Character of America', *The Social Contract* (2000–2001): 91–95, pp. 94–95.

¹⁰⁹ Waldron, 'Dignity and Defamation', p. 1601, n. 14.

¹¹⁰ See note 46 above.

hatred has to do with either the intended or likely effect of speech, meaning that we call something 'incitement to hatred' either when the speech is intended to cause emotions, feelings, or attitudes of hate or hatred or when emotions, feelings, or attitudes of hate or hatred are its likely effect even if the speaker does not set out to incite hatred but nevertheless does so through recklessness or negligence. It goes without saying that the difference between intentional and negligent stirring up hatred may be important as far as issues of moral blame and legal liability are concerned. Some legal regimes may choose to only criminalise intentional stirring up hatred. However, in the present discussion I am primarily interested in articulating the basic ways in which speech can be related to hate or hatred, as part of the project of defining what hate speech is as opposed to justifying how we ought to respond to it. And so I discuss the broad category *the effects of hate speech*, which includes intended as well as negligently created effects.

It is more important to notice that something may count as incitement to hatred quite apart from whether or not the speaker is motivated by hatred or is using speech to express or articulate his or her own hatred. Consider 'Muslims deserve only hatred'. This might be motivated by or express the speaker's hatred, but not necessarily. One could imagine someone engaging in this sort of speech in order to create anti-Muslim feeling in his neighbourhood yet not because he hates Muslims but because he is lonely and wants to be accepted by people who would like him to stir up hatred against Muslims; not in a way that articulates his hatred of Muslims but in a way that symbolises his own self-loathing.

A familiar problem confronts this line of analysis. A good deal of speech concerning members of groups or classes of persons identified by protected characteristics might not be intended to incite hatred and might not actually incite hatred against them, but we might nevertheless think it appropriate to call it 'hate speech'. Suppose a white student goes up to a black student on a university campus and addresses him as 'nigger', without irony and without playfulness or a misguided attempt at friendship. Nobody else is around. This is not intended to, nor does it, incite hatred against black students among other white students. But it is surely hate

speech; and perhaps even a paradigmatic example of hate speech.¹¹¹ Of course, it might be countered that in these sorts of cases the speech could be intended to stir up or have the effect of stirring up *self-hatred* among the objects of speech.¹¹² But just because this can sometimes be the case it does not mean it is always the case. It is equally possible that the intended and achieved purpose is to intimidate the victim, to produce feelings of fear or insecurity, to put him or her off balance, or to change the calculation of whether it is worth going to college.¹¹³ And in that scenario surely it remains plausible to call this ‘hate speech’.

Conversely, much political speech could have the likely effect of stirring up hatred against members of certain groups, but it is not always clear that this speech should count as hate speech. Consider, for example, comments made by the Labour politician, and then Leader of the House of Commons, Jack Straw MP, in an article entitled ‘I Want to Unveil My Views on an Important Issue’, published in the *Lancashire Evening Telegraph* in October 2006, concerning his experiences of talking to Muslim women in his constituency office whilst they are wearing a full veil.

It was not the first time I had conducted an interview with someone in a full veil, but this particular encounter, though very polite and respectful on both sides, got me thinking. In part, this was because of the apparent incongruity between the signals which indicate common bonds — the entirely English accent, the couple’s education (wholly in the UK) — and the fact of the veil. Above all, it was because I felt uncomfortable about talking to someone ‘face-to-face’ who I could not see.¹¹⁴

Some people regarded the article as measured, reasonable, non-derogatory, balanced, informed, respectful in its tone, and thought-provoking. Nevertheless, the *Socialist Worker* reported that the comments led to a new wave of religious hatred against Muslims in the area.¹¹⁵ I do not know whether this report was accurate. But suppose it was. Suppose Straw’s comments unconsciously reinforced the stigmatising assumption that being veiled means being silenced, unwittingly normalised negative stereotypes about Muslim practices

¹¹¹ See, e.g., Matsuda, ‘Public Response to Racist Speech’.

¹¹² *Ibid.*, p. 2336, n. 84.

¹¹³ *Ibid.*

¹¹⁴ Jack Straw, ‘I Want to Unveil My Views on an Important Issue’, *Lancashire Evening Telegraph*, October 5, 2006.

¹¹⁵ Charlie Kimber, ‘Jack Straw’s Veil Comments are Ammunition for Racists’, *Socialist Worker* 2022, October 14, 2006.

in general, implicitly assured people that it was acceptable to hate Muslims, even created new anti-Muslim hatred. What is more, suppose that Straw's comments were always likely to have the effect of stirring up hatred in that part of England, because of existing predispositions to animosity and hostility between Muslim and non-Muslim communities. On the current conceptualisation, Straw's comments were hate speech. But if our intuitions tell us something different, then we may question the veracity of the conceptualisation.

The main thrust of the argument so far has been to challenge the assumption that the best way to explain the meaning of the term 'hate speech' is compositionally, including by drawing on the literal meaning of the term 'hate' consisting of emotions, feelings, or attitudes. But let us now pause to ask this important question: where does the assumption that the word 'hate' in 'hate speech' carries its ordinary or literal meaning come from? Perhaps it comes from what we take to be the semantic constant across a range of other complex terms that include the word 'hate', such as 'hate crime', 'hate incident', 'hate stare', 'hate campaign', 'hate mail', 'hate group', 'hate literature', 'hate propaganda', 'hate movies', 'hate site', 'hate tweets', 'the politics of hate', and 'haters gonna hate'. But the basic problem is that not all compound terms that contain the word 'hate' are partly composed of the ordinary or literal meaning of 'hate'. In other words, it would be wrong to simply assume that the word 'hate' functions semantically within the term 'hate speech' in the same way it does within other complex terms which also contain it.

IV. THE ORDINARY CONCEPT *HATE SPEECH* AS A SEMI- OR QUASI-COMPOSITIONAL CONCEPT?

Nevertheless, before we give up entirely on the compositional approach, there are some alternative analyses of the term that need to be considered because they construe hate speech as having semi- or quasi-compositional semantics. According to these approaches, even if syntactically the term 'hate speech' is a compound noun, semantically its meaning is *not* a function of the literal meanings of the parts that make up the whole term, and understanding the literal meanings of the parts of the whole term may be an unreliable or even highly misleading pathway to understanding the meaning of

the whole term. Nevertheless, the term 'hate speech' is semi- or quasi-compositional in some relevant way.

One approach is to treat 'hate speech' as a partially transparent idiom whose meaning is not directly deducible from the literal meanings of the words 'hate' and 'speech', but whose meaning can be inferred in the sense that the words 'hate' and 'speech' provide a clue to the meaning of the idiom. One possibility is that the word 'hate' is operating as an attributional metaphor, meaning that the word 'hate' refers to something that possesses a quality that can also be attributed to the phenomena referred to by 'hate speech'. The basic idea here is that the word 'hate' could operate in the term 'hate speech' akin to how the word 'zebra' operates in the term 'zebra crossing', namely, to attribute to crossings a quality that is possessed by zebras. The word 'zebra' does not carry its literal meaning into the term 'zebra crossing': namely, as a word that refers to species of African equids with distinctive black and white striped coats. But it does continue to denote *one* of the attributes or qualities possessed by zebras, that is, monochromatic stripiness. In other words, the word 'zebra' in 'zebra crossing' means *like a zebra in at least one respect*. The reference to this quality itself partially contributes or provides a clue to the figurative meaning of 'zebra crossing'.¹¹⁶ In this way the attribute or quality of monochromatic stripiness is mapped from the base domain of zebras to the target domain of crossings. Just as the word 'zebra' does not contribute reference to African equids to the meaning of 'zebra crossing', so the word 'hate' does not contribute reference to hate or hatred to the meaning of 'hate speech'. This is evidenced by the fact, as demonstrated in previous sections, that many instances of hate speech have nothing whatsoever to do with hate or hatred. Nevertheless, maybe the word 'hate' continues to denote at least one of the attributes or qualities of hate or hatred and the reference to this attribute or quality partially contributes or provides a clue to the meaning of the term 'hate speech'. In other words, the word 'hate' in 'hate speech' means *like hate in at least one of its attributes or qualities*.

¹¹⁶ The figurative meaning of 'zebra crossing' is a crossing point for pedestrians marked with black and white stripes, which stands in contrast to its literal meanings, such as a spot where zebras traverse something, the phenomenon of zebras traversing something, or the act of travelling across zebras. Compare this with compound nouns that are more straightforwardly or transparently semantically compositional such as 'pedestrian crossing' and 'hedgehog crossing'.

But which likeness or quality? Perhaps the word 'hate' refers to the attribute or quality of being an emotion or feeling or attitude and contributes this to the meaning of 'hate speech'. Or maybe the word 'hate' refers more specifically to the quality of being a negative emotion or feeling or attitude. Or conceivably it refers to the quality of being mobilising or motivational for human action. Then again, is it really part of the meaning of the term 'hate speech' that the speech in question possesses one or more of the aforementioned qualities? I have my doubts. Even so, I shall not attempt to provide a definitive answer here, safe to say that in each case it would be necessary to show that the relevant quality makes an appearance, without fail, in all *bona fide* instances of hate speech, in the same way that the quality of monochromatic stripiness can be found in all genuine zebra crossings. A far more telling point, however, is that *at most* we can only think of the term 'hate speech' as semi- or quasi-compositional, but even here it would be wrong to assume that the word 'hate' carries with it reference to its most distinctive quality, hate or hatred, that is *intense or extreme dislike, aversion, loathing, antipathy, enmity or hostility toward or against something or someone*. So there is still no escaping the myth of hate even under the proposed analysis.

Alternatively, maybe the term 'hate speech' is a relational metaphor, meaning that it conveys at least one *relational structure* that is common to both feelings or emotions or attitudes of hate or hatred and to the speech in question. In the case of 'pillow talk', for example, the contribution made by the word 'pillow' is not to refer to qualities of pillows but to connote the idea that pillows stand in a certain relation to beds, namely, they are typically found on them, which is the same relation in which pillow talk stands to beds, it happens on or in beds. Perhaps the contribution made by the word 'hate' is also to identify a relational structure that obtains in both hate and hate speech. In other words, the word 'hate' in 'hate speech' means *analogous to hate in at least one of its relational structures*.

But which relation? This depends on the meaning of 'hate' of course, but one possibility is that hate or hatred is typically toward or against something or someone. And so the word 'hate' in 'hate speech' means *toward or against something or someone analogous to how hate is toward or against something or someone*. In its plain or literal sense the word 'hate' does not specify who or what is the object of

hate. But perhaps when it is used in combination with ‘speech’ or ‘crime’, for example, the word ‘hate’ does not convey its plain or literal sense but means something more specific. As Cortese points out, ‘[s]ince the mid-1980s, ‘hate’ began to be used in a much more restricted sense to characterize an individual’s negative beliefs and especially feelings about the members of some other category of people based on their ethnicity, race, gender, sexual orientation, religion, age, or physical or mental disability.’¹¹⁷ This usage persists in current work on hate speech and hate crimes. For example, in her latest article covering these issues Matsuda talks about hate as though it were a target-specific concept.¹¹⁸ Or, to take another example, in her recent book on hate crimes on the Internet, Danielle Keats Citron defines ‘hate’ as saying something negative about or doing something negative to protected groups.¹¹⁹ What this suggests is that the word ‘hate’ can itself have different meanings depending on the context. Its core or standard meaning is *intense or extreme dislike, aversion, loathing, antipathy, enmity or hostility toward or against something or someone*. And this is the meaning it portrays in the term ‘hate mail’ and the slogan ‘haters gonna hate’. But in more specialised contexts, such as in the terms ‘hate speech’ and ‘hate crime’, the word ‘hate’ can mean simply *against members of groups or classes of persons identified by protected characteristics*. This underscores why, under this analysis, the term ‘hate speech’ lacks pure compositional semantics: one cannot deduce the meaning from the literal meaning of the words involved because the term ‘hate’ itself has a figurative or metaphorical meaning. The upshot is that on the current analysis the figurative or metaphorical meaning of ‘hate speech’ is *speech that is toward or against members of groups or classes of persons identified by protected characteristics analogous to how hate is toward or against something or someone*.

Interestingly, in her study of child development and the interpretation of metaphors, Dedre Gentner found that children possess

¹¹⁷ Cortese, *Opposing Hate Speech*, p. 3.

¹¹⁸ Mari Matsuda, ‘Is Peacemaking Unpatriotic? The Function of Homophobia in the Discursive World’, *Journal of Hate Studies* 11 (2013–2014): 9–28, p. 19.

¹¹⁹ Danielle Keats Citron, *Hate Crimes in Cyberspace* (Harvard, MA: Harvard University Press, 2014), p. 16.

an early ability to understand metaphors based on shared object attributes (e.g., 'the tire is like a shoe', *both are made of rubber*) before they develop the ability to interpret metaphors based on shared relational structure (e.g., 'the tire is a shoe', *they both provide a point of connection with the ground*).¹²⁰ Moreover, Gentner has found that adults not merely have an ability, which children lack, to understand relational metaphors but will actually prefer or favour relational metaphors over attributional metaphors in the sense that they will tend to focus primarily on relational commonalities in interpreting metaphors and will rate metaphors based on relational commonalities as more apt than metaphors based on attributional commonalities.¹²¹ Thus, if the term 'hate speech' really does connote certain attributional and relational metaphors, then we cannot expect children to have the ability to understand or interpret 'hate speech' as a relational metaphor, whereas we can expect adults to prefer or favour an interpretation of 'hate speech' as a relational metaphor.

One final approach worth considering treats the terms 'hate speech', 'hate crime', 'hate mail', 'haters gonna hate', and so forth, as what philosophers of language call core-dependent homonyms. The basic idea, due to Aristotle,¹²² is that even though a set of terms which share a common word can have different meanings, understanding their meanings ultimately depends on understanding the core meaning of one of their number, a core meaning on which they are all strongly dependent, whereas understanding the core meaning of the relevant foundational term does not depend on understanding the meanings of the other terms, so there is asymmetry.¹²³ Take, for example, the terms 'healthy body', meaning *in good physical condition*, 'health supplements', meaning *substances designed to provide nutrients that may otherwise not be consumed in sufficient quantities in a normal diet*, and 'healthy urine', meaning *urine that is the colour of pale translucent yellow to gold*. To say these terms are core-dependent

¹²⁰ Dedre Gentner, 'Metaphor as Structure Mapping: The Relational Shift', *Child Development* 59 (1988): 47–59.

¹²¹ *Ibid.*

¹²² Aristotle, *The Metaphysics*, trans. H. Lawson-Tancred (London: Penguin, 1998), pp. 80–81 [1003a–b].

¹²³ Cf. Christopher Shields, *Order in Multiplicity: Homonymy in the Philosophy of Aristotle* (Oxford: Oxford University Press, 1999), ch. 4.

homonyms is to say that even though they mean different things, they each have a strong dependence on a single core meaning upon which they depend. Thus, it might be that the real meaning of the term 'health supplements' is something like *substances designed to provide nutrients that may otherwise not be consumed in sufficient quantities in a normal diet and that preserve good physical condition*. And the real meaning of the term 'healthy urine' is in fact *urine that is the colour of pale translucent yellow to gold and that is an indicator of good physical condition*.

It is not hard to see how the compound terms 'hate speech' and 'hate crime' might be dependent upon, in some cases, the meaning of the term 'hate', namely, *emotions, feelings, or attitudes of intense or extreme dislike, aversion, loathing, antipathy, enmity or hostility towards or against something or someone*. Thus suppose for the sake of argument that 'hate speech' means, in some cases, *negative speech against members of groups or classes of persons identified by protected characteristics*, and 'hate crime' means, in some cases, *criminal acts against members of groups or classes of persons identified by protected characteristics*. Perhaps we could say that in these cases the meanings of 'hate speech' and 'hate crime' depend on the meaning of 'hate' in the following strong ways. The term 'hate speech' means, in some cases, *negative speech against members of groups or classes of persons identified by protected characteristics that expresses the speaker's emotions, feelings, or attitudes of intense or extreme dislike, and so on*. And we could say that the term 'hate crime' means, in some cases, *criminal acts against members of groups or classes of persons identified by protected characteristics that are motivated by the perpetrator's emotions, feelings, or attitudes of intense or extreme dislike, and so on*. But this is by no means always the case, and may not even be typically the case or generally the case. As I have tried to show in this article, there are many instances of hate speech where this putative connection is not merely weak but non-existent. In other words, it is an open question whether the dependency that the putative meanings of the terms 'hate speech' and 'hate crime' have on the core meaning of the term 'emotions, feelings, or attitudes of hate or hatred' are sufficiently common across all meaningful uses of the terms 'hate speech' and 'hate crime' to sustain the idea that these are core-dependent

hyponyms. In other words, it is still possible that the terms 'hate speech' and 'hate crime' can be used in cases where no hate or hatred is involved, and where the meanings of these terms are not dependent on the literal meaning of 'hate'. So long as this is the case we are obliged to take seriously my thesis about the myth of hate.

Notwithstanding everything I have said so far about semantic compositionality and the myth of hate, thus far I have not sought to directly challenge another important assumption found in the literature on hate speech. This is the assumption that the term 'hate speech' is univocal or has a single meaning. Why should we accept this assumption? What if the principal feature of the term 'hate speech' is that it has multiple meanings? How do we analyse it then? Of course, we could simply try to extend some of the approaches employed in this part. We might suppose, for instance, that the term 'hate speech' is doubly figurative or metaphorical; which is to say, the term 'hate speech' is systematically ambiguous between its attributional and relational metaphor.¹²⁴ Nevertheless, once we are no longer bound by the assumption of univocality, we are free to utilise a range of forms of philosophical analysis to get at the semantic multiplicity of the term 'hate speech'. In the second part of this two-part article, 'Family Resemblances', I try to defend this equivocal reading of the term 'hate speech' and to carry out four types of philosophical analysis that can shed some light on the full extent of the family of meanings associated with it.

ACKNOWLEDGMENTS

Versions of this two-part article were presented at the UEA School of Law Seminar in 2013, the UEA School of Politics, Philosophy, Language and Communications Studies internal speaker series in 2016, and the American Philosophical Association (APA) Pacific Division conference in 2016. I am very grateful to all those who attended and provided excellent comments on those occasions. During the course of writing I also benefited immensely from input of ideas and corrections by Susan

¹²⁴ Cf. Gentner, 'Metaphor as Structure Mapping'.

Brison, Eugen Fischer, Michael Frazer, Robin Jeshion, Eric Kaufmann, Mary Kate McGowan, Catherine Rowett, Adriana Sinclair, and Robert Simpson. Finally, I would like to thank the journal's anonymous reviewer for much positive support and numerous helpful suggestions.

OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

*School of Politics, Philosophy,
Language and Communication Studies,
University of East Anglia (UEA),
Norwich Research Park, Norwich, NR4 7TJ, UK
E-mail: alexander.c.brown@uea.ac.uk*