



ALEXANDER BROWN*

WHAT IS HATE SPEECH? PART 2: FAMILY RESEMBLANCES

(Accepted 30 March 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

In ‘What is Hate Speech? Part 1: The Myth of Hate’ (henceforth ‘The Myth of Hate’) I explored the possibility that the ordinary concept *hate speech* is amenable to a form of decompositional conceptual analysis involving three building-block concepts (1) *speech*, (2) *groups or classes of persons identified by protected characteristics*, and (3) *emotions, feelings, or attitudes of hate or hatred*. I then focused on the third

* 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).

building block, hate or hatred. I argued that although there are various possible connections between speech and hate, none appear to capture an essential feature or necessary, defining quality of hate speech. I concluded, therefore, that we should reject what I called the myth of hate. In addition to this, I looked at the following alternative analyses of the term 'hate speech': first, that its semantics are semi- or quasi-compositional, akin to 'zebra crossing' or 'pillow talk'; second, that it is part of a larger group of terms which also contain the word 'hate' and are core-dependent homonyms. But I found that these alternatives did not in fact provide clear-cut reasons for not rejecting the myth of hate.

In this second part I shall proceed with a different working hypothesis to that assumed by the method of conceptual decomposition. I shall build on the premise that the semantics of the term 'hate speech' are non-compositional, so that its meaning is not a function of the literal meanings of its parts. I shall also proceed on the basis that the term 'hate speech' is a relatively opaque idiom, that is, hard to understand simply by looking at the literal meanings of the terms 'hate' and 'speech', and is, therefore, unlike relatively transparent idioms such as 'zebra crossing' or 'pillow talk'.

The main aim of this part, however, is to challenge the assumption that the term 'hate speech' is univocal or has a single meaning and that, all being well, a single shared definition should emerge over time to reflect that single meaning. Instead, I argue that the term 'hate speech' is equivocal, that it denotes a family of meanings, for which there is no one overarching precise definition available.

The remainder of the article is structured thusly. Section II tries to motivate the idea that the term 'hate speech' is an equivocal idiom. Then, Sections III to VI present a detailed analysis of the ordinary concept *hate speech* and ordinary uses of the term 'hate speech' employing four methods of conceptual analysis: *purpose-oriented analysis* (coming to know something of the roles or purposes that we expect the term 'hate speech' to fulfil), *folk platitudes analysis* (coming to know something of how ordinary language users generalise about the term 'hate speech' and, indirectly, about the phenomena it is supposed to refer to), *intuitions about cases analysis* (coming to know something of the intuitions that ordinary language

users have about the application of the term ‘hate speech’ to particular cases), and *ordinary language analysis* (coming to know something of what ordinary language users *do* with the term ‘hate speech’). Section VII responds to the failure of Sections III to VI to formulate a set of necessary and jointly sufficient conditions for the appropriate use of the term ‘hate speech’ by arguing that the concept *hate speech* is what Ludwig Wittgenstein called a ‘family resemblances’ concept.¹ Finally, Section VIII looks at the implications of the foregoing arguments for what we can and should say on the important topic of appropriate and inappropriate ways to combat hate speech.

II. ‘HATE SPEECH’ AS AN EQUIVOCAL IDIOM

In ‘The Myth of Hate’ I argued that ‘hate speech’, despite appearances, is not a relatively transparent idiom like ‘zebra crossing’ and ‘pillow talk’. The term ‘hate’ does not provide the sorts of useful clues to the meaning of ‘hate speech’ which ‘zebra’ provides for ‘zebra crossing’ and ‘pillow’ for ‘pillow talk’. I also believe that ‘hate speech’ is unlike ‘zebra crossing’ and ‘pillow talk’ in another important way. Unlike them, it does not have a single meaning. Of course, the pillow talk of one couple will not be the same as that of a second or third couple. Instances of pillow talk will be varied. But the term ‘pillow talk’ does seem to have a single meaning: *conversation in bed, of an intimate kind, often before or after sex*. Consequently, this term can only be meaningfully applied to finite conjunctions of relatively limited types of context, speech content, emotion or feeling, speaker, and activity. The term ‘hate speech’, by contrast, can be applied to countless permutations of relatively unlimited types of context, speech content, emotions, feelings, or attitudes, speakers, and activity. This suggests strongly to me that we are dealing with multiple meanings.

Some neologisms continue to carry the single meaning intended by those who coined them even when, over time, they become an accepted part of the mainstream lexicon. The term ‘pillow talk’ has to a large extent retained its original meaning as coined. What is more, people who know what the term ‘pillow talk’ means often do

¹ Ludwig Wittgenstein, *Philosophical Investigations*, 4th Edn, P. Hacker and J. Schulte (eds.) (Oxford: Blackwell, 2009), p. 36 [67].

not know, and certainly do not need to know, how and why this term came to possess the meaning it does. They are not familiar with Joseph Dunn's Irish epic tale *Táin Bó Cúalnge*. Interestingly, most ordinary language users who use the term 'hate speech' do so quite comfortably also without having heard of, much less having read, the work of Matsuda. They feel their way into its semantics by seeing how it used by current language users. But, in contrast to 'pillow talk', the term 'hate speech' has acquired a great many more meanings with the passage of time. As coined by Matsuda, the term meant, at its core, *racist insult*. It is now used to mean a bewildering array of things including a great deal that is not racist and not insult.

What I am claiming, in other words, is not simply that a variety of different things can count as hate speech, just as many things can count as pillow talk. I am claiming that the term 'hate speech' has more than one meaning. Of course, it has become something of a cliché to assert that there is profound disagreement about the meaning of the term 'hate speech', disagreement not only among legal scholars and legal professionals² but also among ordinary language users.³ But my claim is not that people disagree about what the correct definition of the term is; after all, that would be consistent with one of the definitions being correct and there actually being a single meaning. Instead, what I am claiming is that the term 'hate speech' is systematically ambiguous; which is to say, it carries a multiplicity of different meanings. On the other hand, I also believe that this multiplicity is more akin to a family than a *mélange* of meanings. I shall say more about this in Section VII.

Of course, as explained in 'The Myth of Hate', some writers have defined the term 'hate speech' in a disjunctive way. And, as such, it could be tempting to think that the term does, after all, have a single meaning albeit a complex disjunctive meaning, similar to how the term 'jade' means *nephrite or jadeite*. On this reading, the term 'hate speech' could mean something like *speech or other expressive conduct which insults or degrades or defames or negatively stereotypes or incites hatred, discrimination or violence against persons or groups of persons*

² See, e.g., Anne Weber, *Manual on Hate Speech* (Strasbourg: Council of Europe, 2009), p. 3.

³ The same point is made by Samuel Walker, *Hate Speech: The History of an American Controversy* (Lincoln, NE: University of Nebraska Press, 1994), p. 8; 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), p. 395.

based on their race or religion or sexual orientation or gender identity or disability and which is intimately connected with feelings, emotions or attitudes of hate or contempt or despisement. The problem with offering a complex disjunctive definition based on lists of phenomena, however, is that no sooner do we think that we have arrived at the definitive lists than we run up against phenomena that most ordinary language users would instinctively want to call 'hate speech' but which are not on the lists. If it is not possible for us to provide definitive lists, and not possible to specify essential qualities shared by all the things on our lists, then we are left with two possibilities. One is that although the term 'hate speech' does have a single meaning at any given time (constituted by the lists), that meaning is constantly changing. A second is that the concept *hate speech* is a systematically ambiguous concept and, therefore, the term 'hate speech' has more than one meaning. I find the first possibility implausible: surely not every new or unconventional application of the term 'hate speech' denotes a change in its meaning.

The best we can hope for, and this is no mean achievement in my view, is to significantly improve our understanding of the ambiguity of the term 'hate speech', by mapping as much of the stunning heterogeneity of the phenomena to which it refers and as much of the vast array of different connotations it carries as possible. The crucial point here is that giving up on the idea that the term 'hate speech' has a single meaning does not necessarily mean accepting forms of analysis that are insubstantive, unrigorous, and uninformative. Quite the opposite.

But what sorts of analyses are appropriate for understanding an equivocal idiom and are at the same time substantive, rigorous and informative? Decompositional conceptual analysis is distinctive partly because it expects to supply a precise definition of the concept or term under analysis, a set of necessary and jointly sufficient conditions for things falling under the concept *hate speech* or being appropriately called 'hate speech'. Unlike decompositional conceptual analysis, however, other forms of analysis embrace the idea that providing such conditions is not the be-all and end-all and may

not even be possible given the nature of the concept, such as when the term in question has multiple meanings or is an umbrella term for a varied collection of phenomena. In particular, my aim is to see whether four other methods of philosophical conceptual analysis, namely, purpose-oriented analysis, folk platitudes analysis, intuitions about cases analysis, and ordinary language analysis, have anything distinctive to say about how we can, and should, characterise (if not define) the ordinary concept *hate speech* and the ordinary term ‘hate speech’ that either has not been or could not be inferred from de-compositional analysis.

Before commencing this work, however, I want to briefly state and then respond to a possible objection to the project I have outlined. The objection is that it is foolish to focus on the methods of analytic philosophy when so many other disciplines have so much to say about the ordinary concept *hate speech*. Psychologists look at hate speech using insights gleaned from their professional experience working with victims of hate speech.⁴ Sociologists have sought to understand the relationship between hate speech and group dynamics.⁵ Linguists have investigated the extent, nature and origins of dehumanising metaphors used in hate speech.⁶ And political scientists and cultural ethnologists have employed the techniques of discourse analysis to assess uses of the term ‘hate speech’ in public and political discourse, including in newspaper articles on politics, in political discussion on the Internet, in political meetings, and in parliamentary debates.⁷ Discourse analysis looks upon the term ‘hate speech’ not as something with a universal, trans-contextual meaning, but as a term that is used by people whose discourse is embedded in particular social practices, psychological states of mind, institutional structures, cultural environments, ideologies, and political hierar-

⁴ See, e.g., Melba Vasquez and Cynthia de las Fuentes, ‘Hate Speech or Freedom of Expression? Balancing Autonomy and Feminist Ethics in a Pluralistic Society’, in M. Brabeck (ed.) *Practicing Feminist Ethics in Psychology* (Washington, DC: American Psychological Association, 2000), p. 226.

⁵ See, e.g., Anthony Cortese, *Opposing Hate Speech* (Westport, CT: Praeger Publishers, 2006), pp. 1–6, 140–142.

⁶ See, e.g., Andreas Musolff, ‘Dehumanizing Metaphors in UK Immigrant Debates in Press and Online Media’, *Journal of Language Aggression & Conflict* 3 (2015): 41–56.

⁷ See, e.g., Shiao-Yun Chiang, ‘“Well, I’m a Lot of Things, But I’m Sure Not a Bigot”: Positive Self-Presentation in Confrontational Discourse On Racism’, *Discourse and Society* 21 (2010): 273–294; David Boromisza-Habashi, *Speaking Hatefully: Culture, Communication, and Political Action in Hungary* (University Park, PA: Penn State University Press, 2013); Katherine Gelber and Luke McNamara, ‘Freedom of Speech and Racial Vilification in Australia: ‘The Bolt Case’ in Public Discourse’, *Australian Journal of Political Science* 48 (2013): 470–484.

chies and conflicts. One must first understand something of these contexts in order to understand the many meanings of the term 'hate speech'.⁸ However, I want to be clear that I intend to utilise the above-mentioned forms of conceptual analysis not because I mean to rule out or deny the usefulness of discourse analysis but because I am interested to learn what they might reveal in addition to anything that discourse analysis might reveal. Moreover, it seems to me that what I call purpose-oriented analysis could capture at least some of what discourse analysis captures. Purpose-oriented analysis tries to tailor our conceptions and characterisations to the roles which people and institutions expect the relevant concepts and terms to play for them. It is alive, then, to the part played by those people and institutions in producing or shaping the contents and semantics of the very concepts and terms they use.

III. PURPOSE-ORIENTED ANALYSIS

In this section I use purpose-oriented analysis to identify possible meanings of the term 'hate speech' which reflect the purposive benefits that usage of the term regularly brings. What are the relevant benefits? Justin C. Fisher counts as a benefit 'anything the person using the concept in question has practical reason to pursue'.⁹ So, for example, Frank Jackson suggests that it may be 'sensible' to seek conceptions of the concepts *free action* and *personal identity* that fit in with the jobs we give these concepts 'in governing what we care about, our personal relations, our social institutions of reward and punishment, and the like'.¹⁰ Likewise, when it comes to identifying and perhaps even favouring different characterisations of the term 'hate speech', maybe the question is whether a given characterisation 'will do any work'.¹¹

⁸ See, e.g., David Boromisza-Habashi, 'Hate speech', *The International Encyclopedia of Language and Social Interaction* (Boston, MA: John Wiley and Sons, 2015), p. 716.

⁹ Justin C. Fisher, 'Meanings and Methodologies', in M. Sprevak and J. Kallestrup (eds.) *New Waves in Philosophy of Mind* (Basingstoke: Palgrave Macmillan, 2014), p. 60.

¹⁰ Frank Jackson, *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Oxford University Press, 1998), p. 45.

¹¹ 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), p. 31.

The sort of purpose-oriented conceptual analysis that I am interested in here maps different meanings of the term 'hate speech' onto different types of work the term is doing or ought to be doing for us. Even if a term performs a beneficial job for those who use it, however, we should also be willing to evaluate (morally speaking) the alleged benefits of the jobs performed by that same term from the perspective of everyone and not simply the users. In other words, it is a matter of mapping meanings onto jobs that are for the good of society as whole or jobs that each person in society has reason to value or care about, objectively speaking. This, of course, makes the conceptual analysis (or 'interpretation') of evaluative concepts itself an evaluative enterprise.¹² Or, as Post puts it, '[w]e must evaluate the status of 'hate speech' so defined in order to determine whether it achieves what we wish to accomplish and whether the harms of the definition will outweigh its advantages.'¹³

Now one of the jobs performed by the term 'hate speech' that I have already touched upon in 'The Myth of Hate' is that it helps legal scholars and journalists to categorise and critically evaluate a body of law that is otherwise heterogeneous. Moreover, legal professionals themselves must define the term 'hate speech' with a view to their particular purposes in regulating speech that ought to be regulated and protected speech that ought not to be regulated.¹⁴ But these are by no means the only or even the most important purposes. No doubt it is possible to conceive of countless jobs that could be performed by the ordinary concept *hate speech* and the non-technocratic term 'hate speech'. But in what follows I shall focus on what seem to be the most obvious or widely posited jobs, including jobs that are broadly beneficial for society as a whole as well as jobs that might be partially beneficial but may also pose a threat to other things we care about, not least free speech. At the same time, I will also endeavour to point out some of the drawbacks with linking the meaning of the term 'hate speech' to these jobs and some of the limitations of purpose-oriented conceptual analysis itself.

¹² Cf. Ronald Dworkin, 'Hart's Postscript and the Character of Political Philosophy', *Oxford Journal of Legal Studies* 24 (2004): 1–37, pp. 5–18; Ronald Dworkin, *Justice for Hedgehogs* (Cambridge, MA: Harvard University Press, 2011), Chap. 8.

¹³ Post, 'Interview', p. 31.

¹⁴ *Ibid.*

Among the many jobs that are, or might be, performed by the concept *hate speech* and the term ‘hate speech’ are: (1) highlighting forms of speech that it is believed disproportionately harm already disadvantaged or victimised members of society,¹⁵ (2) flagging up forms of speech that it is believed either are or have the potential to be very socially divisive or destructive of social cohesion in a diverse, multiracial, multiethnic, multicultural, multi-sexual, multi-gendered, and multi-abled societies,¹⁶ (3) identifying forms of speech that can undermine people’s sense that they are members of society in good standing, who deserve to be treated as equal citizens,¹⁷ (4) providing a means of articulating or giving a particular form and shape to the decisions that societies and legal-political regimes feel they need to make, whether explicitly or implicitly, about forms of publically acceptable speech, the appropriate tone of public debate and, more generally, the imposition of civility norms,¹⁸ (5) labelling forms of speech that it is believed may run contrary to fundamental democratic values or even to political legitimacy itself.¹⁹

Of course, if we think that our characterisations of the ordinary concept *hate speech* should be tailored to one rather than another of the aforementioned jobs, then our characterisations will tend to focus on some rather than other aspects of hate speech. So, for

¹⁵ See, e.g., Mari Matsuda, ‘Public Response to Racist Speech: Considering the Victim’s Story’, *Michigan Law Review* 87 (1989): 2320–2381; Charles Lawrence III, ‘If He Hollers Let Him Go: Regulating Racist Speech on Campus’, *Duke Law Journal* (1990): 431–483; 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); Richard Delgado, ‘Toward a Legal Realist View of the First Amendment’, *Harvard Law Review* 113 (2000): 778–802; Alexander Brown, *Hate Speech Law: A Philosophical Examination* (London: Routledge, 2015), Chap. 3.

¹⁶ 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. 129; David Brink, ‘Millian Principles, Freedom of Expression, and Hate Speech’, *Legal Theory* 7 (2001): 119–157, p. 119; Bhikhu Parekh, ‘Hate Speech: Is There a Case for Banning?’, *Public Policy Research* 12 (2005–2006): 213–223, p. 223; Brown, *Hate Speech Law*, Chap. 6.

¹⁷ See, e.g., Jeremy Waldron, ‘Dignity and Defamation: The Visibility of Hate’, *Harvard Law Review* 123 (2010): 1596–1657, pp. 1621–1623; Brown, *Hate Speech Law*, Chap. 5.

¹⁸ See, e.g., Judith Butler, *Excitable Speech: A Politics of the Performative* (New York, NY: Routledge, 1997), p. 77; Robert Post, ‘Hate Speech’, in I. Hare and J. Weinstein (eds.) *Extreme Speech and Democracy* (Oxford: Oxford University Press, 2009), pp. 127–136.

¹⁹ See, e.g., Charles Lawrence III, ‘Cross Burning and the Sound of Silence: Anti-Subordination Theory and the First Amendment’, *Villanova Law Review* 37 (1992): 787–804, pp. 792, 800; Erik Bleich, *The Freedom to Be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism* (Oxford: Oxford University Press, 2011), pp. 12–13; Brown, *Hate Speech Law*, Chap. 7. Alexander Brown, ‘Hate Speech Laws, Legitimacy, and Precaution: A Reply to James Weinstein’, *Constitutional Commentary*, forthcoming. Cf. 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.

example, when critical race theorists use the term ‘racist hate speech’ with a view to emphasising the harmful effects of everyday racist speech on members of historically oppressed or victimised groups, this leads to a conception of hate speech that foregrounds the use of insults, slurs, epithets, and also threatening signs and symbols, particularly in face-to-face encounters.²⁰ To take another example, it is because Waldron has a core interest in the role the term ‘hate speech’ might play in identifying forms of speech that threaten people’s sense that they are members of society in good standing that he feels compelled to conceive of hate speech in terms of group libel.²¹ Finally, James Weinstein and Ivan Hare are interested in the job that the term ‘hate speech’ and related terms – such as ‘extreme speech’ – can perform in identifying forms of speech ‘that many believe pose an unacceptable threat to essential values in modern multicultural democracies, or in some cases, to democracy itself.’²² This in turn means that they characterise hate speech as ‘inciting others to hatred of the target group or seeking to encourage the audience to discriminate against them’.²³ Now these analyses might be well-motivated in virtue of the assumptions made about the respective purposes served by the term ‘hate speech’, but this does not mean they cannot be criticised. A conception of hate speech might be criticised as being under- or over-inclusive given the job it is designed to do, for example.²⁴

A much broader job performed by the term ‘hate speech’ is simply to express disapproval of certain forms of speech. It is not merely an evaluative term but also a pejorative term in that sense. When we describe something as ‘hate speech’ we intend this to be a form of condemnation, invariably of the speech but also sometimes of the speaker.²⁵ It is a kind of speech that deserves *censure*. In order

²⁰ See, e.g., Matsuda, ‘Public Response to Racist Speech’, pp. 2332, 2335, 2358; Lawrence, ‘If He Hollers Let Him Go’, p. 452; Lawrence, ‘Cross Burning and the Sound of Silence’, pp. 787–788; Lawrence et al., ‘Introduction’, p. 1; Delgado, ‘Toward a Legal Realist View of the First Amendment’, pp. 786–789.

²¹ See, e.g., Waldron, ‘Dignity and Defamation’, pp. 1597, 1600–1601.

²² Weinstein and Hare, ‘General Introduction’, p. 4.

²³ *Ibid.*

²⁴ For a challenge to the narrowness of Waldron’s characterisation of hate speech, see Brown, *Hate Speech Law*, Chap. 5.

²⁵ See, e.g., Boromisza-Habashi, ‘Hate speech’, p. 715.

for something to count as hate speech given the aforementioned purpose, therefore, it must be the sort of thing that is worthy of disapproval or condemnation; otherwise the concept is not doing one of the jobs that we expect it to do. Of course, the job of expressing disapproval of certain forms of speech is performed by many terms besides ‘hate speech’. Think of ‘unjust speech’, ‘dangerous speech’, ‘ill-considered speech’, and ‘harmful speech’. Then again, the term ‘hate speech’ seems to be particularly powerful in the present climate. We tend to reserve the term for forms of speech that we have special reason to be intolerant of within diverse societies. This perhaps explains why people engaged in hotly contested debates surrounding institutional racism, religious intolerance, sexual and transgender politics, national identity and immigration, or disability rights, to name but a few, sometimes resort to accusing the other side of engaging in ‘hate speech’. Consider the arrests and subsequent acquittals of the Kenyan politicians Moses Kuria and Junet Mohamed on suspicion of inciting hatred, both of whom claimed innocence but also accused the other of guilt.²⁶

However, one danger with linking the meaning of the term ‘hate speech’ to the job of expressing moral disapproval is that some users will use the term to refer to forms of speech that perhaps only they and few others believe are worthy of disapproval or condemnation. What is more, dominant sections of society or in some instances undemocratic, authoritarian and oppressive governments might exploit the term ‘hate speech’ not merely to condemn the speech of others but to discredit or silence the speech of people with whom they simply disagree. They use the term ‘hate speech’ not merely to censure but also to *cancel*. If the term ‘hate speech’ is being used by such groups to perform these jobs, the result may be a significant disbenefit for society as a whole, in terms of chilling or restricting legitimate forms of self-expression, public discourse, and dissent.²⁷ By way of illustration, some notable civil libertarian scholars in the United States have suggested that the term ‘hate speech’ is being used by governmental authorities and liberal elites to enforce a

²⁶ Unkown, ‘MPs Moses Kuria and Waititu Acquitted of Hate Speech Charges’, *Capital News*, February 20, 2017. Available at www.capitalfm.co.ke/news/2017/02/mps-moses-kuria-waititu-acquitted-hate-speech-charges/.

²⁷ See, e.g., Boromisza-Habashi, *Speaking Hatefully*; Kenan Malik, ‘Interview’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012).

particular type of ideology that prizes political correctness over freedom of expression.²⁸ Much the same suggestion has been made by some writers in Europe who have been publically accused of engaging in hate speech or who have been prosecuted for, or fear that they might be prosecuted for, engaging in hate speech. Writing in the left-wing Hungarian newspaper, *Népszabadság*, in 1996, for example, the Hungarian revisionist historian Mária Schmidt claimed that the meaning of the term ‘hate speech’ ‘can perhaps only be compared to concepts like class enemy or enemy of the people, and its construction is downright Orwellian’.²⁹ In a similar vein, consider this passage taken from an article posted on a Holocaust denial website by the anti-Zionist activist Michael Rivero. ‘These phrases, ‘hate speech’ and ‘anti-Semite’, are well-worn devices to shut up a critic of Israel without having to answer the criticisms.’³⁰ Likewise, many on the Christian right in the United States affirm that the term ‘hate speech’ is synonymous with a liberal attack on literalist interpretations of the Bible’s teaching on homosexuality. In a recent interview with CBN News the Republican Senator Marco Rubio declared, ‘If you think about it, we are at the water’s edge of the argument that mainstream Christian teaching is hate speech.’³¹ A note posted on the website of the Westboro Baptist Church puts a similar point even more forcefully. ‘Just because you rage against God and make laws that say you cannot use ‘hate speech’ (a/k/a – you may not speak of the Bible standards) in the UK does NOT mean you will not get the message that God Almighty intends for you to get.’³² Then again, it seems equally likely that the story being told by some civil libertarians about how governmental authorities and liberal elites have been using the concept *hate speech* to silence valuable speech and precious forms of dissent are themselves using the term ‘so-called hate speech’ to do a job the social value of which

²⁸ See, e.g., Henry Louis Gates, ‘War of Words: Critical Race Theory and the First Amendment’, 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), pp. 17–18; 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), p. 73.

²⁹ Cited in Boromisza-Habashi, *Speaking Hatefully*, p. 84.

³⁰ Available at <http://whatreallyhappened.com/WRHARTICLES/hatespeech.html>.

³¹ David Brody, ‘Rubio Warns of ‘Clear, Present Danger’ to Christianity’, *CBN*, May 26, 2015. Available at www.cbn.com/cbnnews/politics/2015/May/Rubio-Warns-of-Clear-Present-Danger-to-Christianity/.

³² Available at www.godhatesfags.com.

can also be called into question, namely, the job of defending indefensible forms of speech. And maybe the claim that university campus codes designed to combat racist hate speech are political correctness gone mad is itself, as John K. Wilson puts it, part of a 'melodrama' staged by conservatives to attack codes they simply do not like.³³ And perhaps the fundamentalist or evangelical Christian view that laws banning incitement to hatred on grounds of sexual orientation are a form of religious persecution are themselves an indirect attempt to incite or justify discrimination on grounds of sexual orientation.

This means that if we want to get at the meaning of the ordinary concept *hate speech* through the use of purpose-oriented conceptual analyses, then tough choices will have to be made about what jobs or purposes can be deemed relevant (because beneficial) and, therefore, what things the search for meaning ought to capture or take its lead from.³⁴ Assuming that the aforementioned allegations and counter-allegations about the types of malign jobs which the term 'hate speech' is being used to perform are accurate, then perhaps we should not tailor our characterisations of the meaning of that term to *all* jobs, and certainly not to the job of ideologically-motivated censorship, but instead to jobs that are objectively beneficial for society as a whole.

At any rate, one upshot of appealing to this sort of analysis is that we are likely to end up with more than one appropriate purpose or function for the concept *hate speech*. What is more, a plurality of purposes or functions is likely to produce a plurality of meanings. Even if we just focus on (1) highlighting forms of harmful speech, (2) flagging up socially divisive forms of speech, (3) identifying forms of speech that can undermine people's sense of equality, (4) articulating civility norms, and (5) labelling up forms of speech that undermine democracy, we are likely to end up with a family of different meanings of the term 'hate speech', each slightly different to the next. My own view, to be articulated more fully in Section VII, is that we should embrace rather than resist this outcome, and thereby jettison the assumption that if we only look hard enough we can find only one purpose and a single definition.

³³ John K. Wilson, *The Myth of Political Correctness: The Conservative Attack on Higher Education* (Durham, NC: Duke University Press, 1995), p. 91.

³⁴ Cf. Fisher, 'Meanings and Methodologies', p. 60.

IV. FOLK PLATITUDES ANALYSIS

Folk platitudes analysis, as I shall mean it, is the identification and examination of commonplace generalisations or folk platitudes held about the term or concept in question.³⁵ Some folk platitudes about hate speech are shared across many different countries or socio-legal communities, whereas others are idiosyncratic to just one or two countries or cultural communities. The method takes these folk platitudes as things that ought to be taken seriously, but it also recognises that these folk platitudes can be mistaken. They tend to be affirmed by people before they have been exposed to more detailed and rigorous thinking about the nature of hate speech, including the method of comparing and contrasting different platitudes, and before they have been invited to think about more unusual examples or unusual ways of thinking about run of the mill examples.

What, if any, folk platitudes exist about the ordinary concept *hate speech*? In what follows I shall examine four such platitudes. The first is that 'hate speech' is an evaluative term in the sense that one cannot, at least with sincerity or without irony, label something 'hate speech' without thereby passing a negative normative judgement upon it, in the same way one cannot call something 'unjust' without thereby claiming it to be a bad thing. That the term 'hate speech' is negatively evaluative is also suggested by the accompanying language that people use to refer to the extent of hate speech or to increasing levels of hate speech or to the way hate speech is spread. It is not uncommon for people, including academics, to refer to an 'epidemic'³⁶ or 'contagion'³⁷ of hate speech. If the term 'hate speech' did not signal something that was morally bad or at least unwelcome, then such metaphors would seem rather incoherent. The fact that they are not jarring or peculiar sounding is telling. After all, we rarely, if ever, talk of epidemics or contagions of apple pie. But can it really be the case that all uses of the term 'hate speech' imply a negative evaluation of the speech in question? Do not people – critics of hate speech law in particular – sometimes use locutions such as

³⁵ *Ibid.*, p. 63.

³⁶ See, e.g., Matsuda, 'Public Response to Racist Speech', pp. 2332, 2370.

³⁷ See, e.g., Stephen L. Newman, 'Should Hate Speech be Allowed on the Internet? A Reply to Raphael Cohen-Almagor', *Amsterdam Law Forum* 2 (2010): 119–123, p. 120.

‘so-called hate speech’ or ‘what is commonly referred to as ‘hate speech’’ when talking about forms of speech that they believe are not entirely a bad thing or at least not bad enough to warrant legal prohibition or that they believe could be valuable in certain ways? They might say, ‘So-called hate speech serves an important function in society by allowing people to let off steam verbally rather than engaging in acts of violence and, what is more, provides the rest of society with a window into the minds of racists which generally speaking is a good thing if we want to know how to tackle racism.’ Or, ‘What is commonly referred to as ‘hate speech’ in fact includes various forms of artistic, political, religious and dissenting speech – all of which is vitally important in maintaining vibrant and diverse public discourse.’ On the other hand, arguably even in such cases people are implicitly acknowledging the fact that the ordinary, non-ironic meaning of the term ‘hate speech’ involves a negative connotation. It is this accepted connotation that they are seeking to challenge in some way.

A second, much less straightforward folk platitude is that hate speech is about certain types of people. Ordinarily we would not, I think, use the term ‘hate speech’ to describe the words of someone who expresses loathing of mosquitoes, who makes false claims about the cleanliness of rats, or who spreads negative stereotypes about vultures, for instance. By the same token, we tend not to count as hate speech generalised misanthropic speech that denigrates or vilifies the entire human race.³⁸ Nevertheless, the mere fact that speech is about or directed at particular human beings is not enough to make it hate speech. The concept *fighting talk* also pertains to speech directed at human beings only, yet not all fighting talk is hate speech. It seems that genuine hate speech has to do with members of groups or classes of persons identified by only certain characteristics. But which ones? Here the platitude does not provide a definitive answer. If pressed, language users could provide lists of characteristics, and perhaps many of these lists would include race, ethnicity, religion, sexual orientation, gender identity, and disability. But it is uncertain whether such lists would define the limits of the concept. Take the *South Park* episode from 2005 in which Cartman’s school presentation about ‘ginger kids’ – who it is claimed are suffering from ‘gin-

³⁸ Parekh, ‘Hate Speech’, p. 214.

gervitis' and have 'no souls' – causes Kyle to object, 'That's not a lecture, it's a hate speech.'³⁹ This speech is insulting or disparaging about people with a particular type of hair colour, but we do not, I think, balk at Kyle's usage. It is not entirely clear who we are supposed to be laughing at here, but even if there is something humorous about the scene, the use of the term 'hate speech' does not seem unintelligible. Perhaps it makes sense to us because hair colour is being used as a proxy for some form of racial or ethnic group identity. Then again, what if it is exactly what it appears to be, a hate speech about people with ginger coloured hair? This example would then suggest that the concept is applicable to more characteristics than our putatively definitive list had assumed.⁴⁰

Consider two other hard cases for this platitude. First, think about cases where someone expressly denies that a certain group of people are human beings. Consider this utterance (which has been attributed to Adolf Hitler): 'The Jews are undoubtedly a race, but they are not human.' Is this a counter-example to the folk platitude that hate speech is always about or against human beings? I would argue not. The folk platitude belongs to and is about the actual world, a world in which Jews are human, as opposed to a possible world in which Jews are not human. The key issue is that the folk platitude holds true in the actual world, so whether or not it holds true in a possible world in which Jews are not human is moot. Here in the actual world the utterance is about human beings and it is hate speech. Of course, there may be people who believe that Jews are not human. And no doubt such people might also believe that the above utterance is *not* hate speech. But I simply take it as read that such people are mistaken on both counts.

Another hard case is speech that denigrates a set of religious beliefs and practices or that defiles the reputation of someone who embodied those beliefs and practices such as a religious prophet. Consider Salman Rushdie's 1988 book *The Satanic Verses* or Kurt Westergaard's cartoons of the Prophet Muhammad published in the

³⁹ *South Park*, Season 9, Episode 11, aired on *Comedy Central* US, November 9, 2005.

⁴⁰ 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.

Danish newspaper *Jyllands-Posten* in 2005. Ostensibly both examples of speech are about a religion and its prophet rather than about a group or class of persons identified by the protected characteristic of religion. And so there may be a tendency to want to not count them as cases of hate speech.⁴¹ Then again, it could be that the semantics of such speech are deeper or more complex than they at first appear. Maybe *some* statements about or caricatured depictions of the Prophet Muhammad actually are about all Muslims or at least are of Muslims in general, and amount to negative stereotypes, stigmatisation, or even defamation.⁴² As such, they can be counted as hate speech. Thus it is feasible that a cartoon depicting the Prophet Muhammad as a terrorist or suicide bomber has something close to the following semantic content, 'Muslims follow the Quran, a book that commands them to emulate Muhammad, but he was a violent, vengeful man, who waged wars against his neighbours, terrorised poets, and beat his own bride, so it is no surprise that all Muslims throughout the world continue to behave in these ways.'

Turning to a third folk platitude that also needs to be handled with great caution, of those ordinary people who have come across the term 'hate speech' before, or are being confronted with it for the first time, I suspect that many would simply assume that hate speech necessarily has something to do with emotions, feelings, or attitudes of hate or hatred. In 'The Myth of Hate' I put forward numerous counter-examples: examples of what we might intuitively call 'hate speech' that lacked this essential involvement with or necessary connection to hate or hatred.

Perhaps a large part of what fuels this myth is that *some* forms of hate speech *are* intimately connected to or bound up with emotions, feelings, or attitudes of hate or hatred. For example, in the case of incitement to hatred the word 'hatred' more likely than not does carry its ordinary or literal meaning. Indeed, what is so troubling

⁴¹ See, e.g., Parekh, 'Hate Speech', p. 215; Robert Post, 'Religion and Freedom of Speech: Portraits of Muhammad', *Constellations* 12 (2007): 72–90, p. 84; Waldron, 'Dignity and Defamation', pp. 1612–1613.

⁴² See, e.g., Tariq Modood, 'Muslims, Incitement to Hatred and the Law', in J. Horton (ed.) *Liberalism, Multiculturalism, and Toleration* (Basingstoke: Macmillan, 1993), p. 145; Tariq Modood, 'The Liberal Dilemma: Integration or Vilification?', *International Migration* 44 (2006): 4–7, p. 4; Tariq Modood, 'Hate Speech: The Feelings and Beliefs of the Hated', *Contemporary Political Theory* 13 (2014): 104–109, pp. 106–108; Glyn Morgan, 'Mill's Liberalism, Security, and Group Defamation', in G. Newey (ed.) *Freedom of Expression: Counting the Costs* (Cambridge: Cambridge Scholars Press, 2007), pp. 128–129; Erik Bleich, 'On Democratic Integration and Free Speech: Response to Tariq Modood and Randall Hansen', *International Migration* 44 (2006): 17–22, p. 21.

about incitement to hatred is precisely that it does aim or is likely to stir up or awaken emotions, feelings, or attitudes of intense or extreme dislike, aversion, loathing, antipathy, enmity or hostility of the sort that are liable to spill over into acts of discrimination or violence.⁴³ But incitement to hatred is only one among many different forms of hate speech. As I shall try to show in Section V, not everything that is intuitively labelled 'hate speech' amounts to incitement to hatred.

I turn finally to the folk platitude that hate speech is a type of regulatable speech. This is to think that for the most part hate speech *may* be regulated, not that it *should* be regulated all things considered. Regulatable speech is contrasted with speech which for the most part may *not* be regulated because it enjoys special protection under a constitutional right to freedom of expression. I do not mean to claim that this folk platitude is affirmed in every country or socio-legal community on the planet. Instead, my claim is that it is probably affirmed in Canada, in many parts of Europe, in Australia, in parts of Africa, in parts of South America, in parts of Asia and East Asia, and in some parts of the Middle East. As for exceptions, it is tempting to assume that Americans share a civil libertarian interpretation of United States Supreme Court jurisprudence that says hate speech is a protected category of speech under the First Amendment and therefore not a type of regulatable speech constitutionally speaking.⁴⁴ But this assumption might be premature. For one thing, there are high-profile shapers of public opinion in the United States, such as the CNN anchor Chris Cuomo, who have publicly declared that hate speech is not protected speech under the First Amendment. Another, admittedly crude, measure of whether ordinary Americans think that hate speech is regulatable speech is whether or not they think it should be regulated. The opinion pollsters Rasmussen Reports and YouGov have both recently published online polls about Americans' attitudes toward the criminali-

⁴³ Cf. Alexander Brown, 'The Racial and Religious Hatred Act 2006: A Millian Response', *Critical Review of International Social and Political Philosophy* 11 (2008): 1–24, pp. 5, 13–14; Brown, *Hate Speech Law*, pp. 66–71.

⁴⁴ See Floyd Abrams, 'On American Hate Speech Law', in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012). Cf. Brown, *Hate Speech Law*, Chap. 2.

sation of hate speech.⁴⁵ These polls indicated low levels of ‘I don’t know’ responses and even revealed that a small majority support hate speech laws. Likewise, in 2017 Yougov United Kingdom undertook a poll of British attitudes to incitement to hatred laws. This poll revealed strong majority support for such laws.⁴⁶ These results would be highly surprising if people thought that hate speech was nonregulatable.

Arguably the folk platitude that hate speech is a type of regulatable speech is consistent with or might even be supported by our beliefs about the sorts of jobs that we expect the ordinary concept *hate speech* to perform. If we believe, for example, that highlighting forms of speech that disproportionately harm already disadvantaged or victimised members of society is an important job or role of the ordinary concept *hate speech*, then we may be more inclined to accept that hate speech is a type of regulatable speech. Indeed, when people who affirm this folk platitude are asked the question, ‘What is it about hate speech that makes it regulatable?’, one likely answer is another folk platitude, that hate speech causes harm of one kind or another, directly or indirectly.⁴⁷ Does this mean that something can be counted as hate speech only if it actually causes harm? Not necessarily. The folk platitude is that hate speech has the *tendency* to cause harm – this tendency is one of its features – not that it always causes harm. And so ordinary language users would, and should, say ‘B was the victim of hate speech and is so often the case he was harmed by it,’ but they would not normally say, ‘B was the victim of

⁴⁵ See Rasmussen Reports, ‘31% Favor Ban on Hate Speech’, June 6, 2013. Available at www.rasmussenreports.com/public_content/politics/general_politics/june_2013/31_favor_ban_on_hate_speech. YouGov Unites States, ‘America Divided on Hate Speech Laws’, October 2, 2014. Available at <https://today.yougov.com/news/2014/10/02/america-divided-hate-speech-laws/>. And YouGov United States, ‘Half of Democrats Support a Ban on Hate Speech’, May 20, 2015. Available at <https://today.yougov.com/news/2015/05/20/hate-speech/>.

⁴⁶ Cited in Alexander Brown, ‘New Evidence Shows Public Supports Banning Hate Speech Against People With Disabilities’, *The Conversation*, March 1, 2017. Available at <https://theconversation.com/new-evidence-shows-public-supports-banning-hate-speech-against-people-with-disabilities-73807>.

⁴⁷ See, e.g., Delgado, ‘Toward a Legal Realist View of the First Amendment’, pp. 787–788; Alexander Tsesis, *Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements* (New York, NY: New York University Press, 2002), p. 211, n. 1; Richard Delgado and Jean Stefancic, *Understanding Words That Wound* (Boulder, CO: Westview Press, 2004), Chap. 1; Brown, *Hate Speech Law*, Chap. 3.

hate speech, but, as is so often the case, it left him unharmed.’ Note, this platitude is weaker than the claim hate speech *necessarily* causes harm but is stronger than the claim that hate speech is merely *capable* of causing harm.⁴⁸

Returning to the folk platitude about regulatability, if this platitude is true, then in order for something to count as hate speech it must be something that we regard as regulatable. Although this folk platitude may well serve some useful purposes, I also believe that accepting it may have some potentially undesirable consequences for how we use the term ‘hate speech’, particularly in countries or socio-legal communities that impose a high threshold on regulatability. Focusing on hate speech as a type of regulatable speech may, for example, push some countries or socio-legal communities in the direction of a narrow conception of the ordinary (as well as the legal) concept *hate speech* according to which something is hate speech only if it amounts to incitement to hatred, discrimination or violence. This reflects two points. The first, already touched upon, is the importance placed on the right to freedom of expression in many legal regimes. This creates a propensity not to regard speech as regulatable unless it is a special case or exceptional in some way. The second point is that when the term ‘hate speech’ is put to legal purposes, especially in restricting speech, there is a need for ‘definitional refinement’⁴⁹ so that the term might be operationalised or justicised. This technical requirement might also lead some countries or socio-legal communities to reject or discount as proper instances of hate speech forms of speech that are difficult to define in statutes, not easily understood and implemented by legal professionals, including the judiciary, the police, and public prosecutors, and hard to comprehend or grasp by ordinary members of the public who must be able to reasonably foresee the scope of speech restrictions. Thus, in countries or socio-legal communities that impose a relatively high threshold on regulatability, accepting the folk platitude that hate speech is a type of regulatable speech may restrict the scope of the ordinary concept *hate speech* to a narrow range of speech, which may in turn undermine important jobs that anyone might

⁴⁸ Cf. Katherine Gelber, ‘Reconceptualizing Counter-speech in Hate Speech Policy (with a Focus on Australia)’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012), p. 213.

⁴⁹ Tarlach McGonagle, ‘A Survey and Critical Analysis of Council of Europe Strategies for Countering “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), p. 457.

reasonably expect the concept to do, such as to identify or flag up forms of speech that disproportionately harm already disadvantaged or victimised members of society.

What all of the above demonstrates, I think, is the need for systematic thinking about folk platitudes: placing our folk platitudes about ordinary concepts *hate speech* into some kind of coherent or ordered system along with our beliefs about the important jobs that we expect those same concepts to do, and of course along with our linguistic intuitions about cases. Aside from what I have already done to challenge the folk platitude that hate speech necessarily has something to do with emotions, feelings, or attitudes of hate or hatred, I do not attempt here to provide a fully-worked out systematisation: I confine myself to sketching out the various elements that it should seek to systematise.

V. INTUITIONS ABOUT CASES ANALYSIS

Intuitions about cases analysis, as I shall understand it, involves examining intuitions about whether given cases (actual and possible) are or are not cases of hate speech primarily via linguistic intuitions about whether using the term 'hate speech' to describe given cases (actual and possible) either seems quite obviously appropriate or clearly inappropriate. This approach taps into the fact that when faced with the question, 'What is hate speech?', competent users may simply say, 'I might not be able to define it exactly, but I know it when I hear it or read it or see it.' The relevant intuitions need not be universally held, but they must be widespread among people who use the concepts, that is, among ordinary users of the term 'hate speech'. Ostensibly, the aim of the analysis is to produce conceptions or meanings that fit with or preserve the truth of intuitions about cases. However, although these intuitions are treated as initially credible, they are not taken to be infallible, meaning that they can also be shown to be mistaken based on countervailing evidence.

Before I begin I also wish to address two possible objections to this form of analysis. The first is that it is difficult to get the analysis off the ground for the simple reason that 'hate speech' is a technocratic term which ordinary people rarely use (so the objection goes). In other words, there is not a stock of standard uses to draw on for the purposes of triggering linguistic intuitions because this is not a

mainstream term. However, I reject the premise of the argument. At the start of 'The Myth of Hate' I provided evidence of how the term 'hate speech' has entered into ordinary language, mainstream media, and popular culture. It is no longer merely a legalistic or technocratic term. Moreover, even if ordinary language users do not hear or read or use the term 'hate speech' on a daily basis, they may nevertheless have some grasp of its meanings and may have intuitions, even if implicit, about whether or not it is appropriately applied to given cases simply based on the fact that they use it now and then.

A second possible objection is that the method of looking at ordinary language is irredeemably conservative. Insofar as the method looks at, as Herbert Marcuse puts it, the 'totalitarian scope of the established universe of discourse',⁵⁰ it is bound to reinforce not only the linguistic practices of the society in question – in the present case practices surrounding the use of the term 'hate speech' – but also the wider array of social, political, and legal practices that support, and are supported by, those linguistic practices – such as the practice of prohibiting or not prohibiting hate speech. However, as Alan Wertheimer has argued, there is no necessary connection between documenting and identifying patterns (and sometimes confusions) within standard everyday uses of language, on the one hand, and accepting or justifying those linguistic practices, much less accepting or justifying the wider array of social, political, and legal practices that support, and are supported by, those linguistic practices, on the other hand.⁵¹ Indeed, according to Marcuse himself, linguistic analysis could support reform of the status quo if particular attention is paid to 'ordinary language in really controversial areas', and if the ambition is to uncover 'muddled thinking where it seems to be the least muddled, uncovering the falsehood in so much normal and clear usage'.⁵² Surely if this applies to anything it applies to analysis of the term 'hate speech', including using intuitions about cases to challenge folk platitudes such as that hate speech necessarily has something to do with emotions, feelings, or attitudes of hate or hatred.

Now for the analysis itself. There are, I think, *at least* five broad types of speech that intuitively fall under the ordinary concept *hate*

⁵⁰ Herbert Marcuse, *One-Dimensional Man* (Boston: Beacon Press, 1964), p. 195.

⁵¹ Alan Wertheimer, 'Is Ordinary Language Analysis Conservative?', *Political Theory* 4 (1976): 405–422, p. 410.

⁵² Marcuse, *One-Dimensional Man*, p. 195.

speech and that are plausibly called ‘hate speech’. Before setting out the types, however, I first need to make some clarifications. The five broad types of hate speech are not intended to be mutually exclusive: some forms of speech fall into more than one of the types. In addition, I do not pretend that this list of five types is exhaustive such that something is hate speech only if it is an instance of one or more of these types. There may well be other types. Indeed, it is quite possible that these broad types could be carved up differently, creating a much longer list potentially. In addition, I am not suggesting that these types all share some common quality or universal feature over and above the quality of being hate speech. So nothing is being assumed here about the essence of hate speech. Rather, the purpose of setting out the different types is simply to illuminate something of the equivocal semantics of the term ‘hate speech’.

First, I suspect that most users of the term ‘hate speech’ would, if asked, apply the term to insults, slurs, epithets, ethnophaulisms, antilocutions, or other words of disparaging abuse or vilification targeted at members of groups or classes of persons identified by protected characteristics. Consider people who direct ethnophaulisms like ‘black bastard’, homophobic slurs like ‘dirty faggot’, or xenophobic epithets such as ‘cockroach’ or ‘bogus asylum-seeking’ at other people whom they perceive to possess certain characteristics.⁵³ Legal scholars,⁵⁴ social psychologists,⁵⁵ linguists,⁵⁶ and philosophers⁵⁷ alike have identified such speech as paradigmatic hate speech. I do not have space here to discuss in detail different theories of the semantics of slurs and epithets, but I shall briefly mention three leading accounts. On the first, slurs express, encode, implicate

⁵³ Countless other examples could be mentioned, of course. Indeed, the sociolinguist and lexicographer Irving Allen classifies ethnic or racial slurs into six categories: physical traits, personal traits, personal names, food habits, group names, and other (miscellaneous). See his *The Language of Ethnic Conflict: Social Organization and Lexical Culture* (New York, NY: Columbia University Press, 1983).

⁵⁴ See, e.g., Matsuda, ‘Public Response to Racist Speech’, p. 2332; Kent Greenawalt, *Fighting Words: Individuals, Communities and Liberties of Speech* (Princeton, NJ: Princeton University Press, 1995), p. 47; Delgado, ‘Toward a Legal Realist View of the First Amendment’, p. 786; Steven J. Heyman, *Free Speech and Human Dignity* (New Haven, CT: Yale University Press, 2008), p. 165; Douglas M. Fraleigh and Joseph S. Tuman, *Freedom of Expression in the Marketplace of Ideas* (Thousand Oaks, CA: Sage, 2011), p. 139.

⁵⁵ See, e.g., Brian Mullen and Tirza Leader, ‘Linguistic Factors: Antilocutions, Ethnonyms, Ethnophaulisms, and Other Varieties of Hate Speech’, in J. F. Dovidio et al. (eds.) *On the Nature of Prejudice* (Oxford: Blackwell, 2005), p. 192.

⁵⁶ See, e.g., Geoffrey Hughes, *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World* (London: Routledge, 2006), p. 220.

⁵⁷ See, e.g., Waldron, ‘Dignity and Defamation’, p. 1600.

or connote derogatory ideas or negative stereotypes about or pertaining to the subject of the slur.⁵⁸ So 'Black bastard' might have the semantic or locutionary content, 'lazy, stupid, aggressive, criminal good for nothing black person,' whereas 'dirty fagot' could mean 'effeminate, promiscuous, sexually confused, sexually deviant, dangerous, HIV-carrying gay person' and 'bogus asylum-seeking' could connote 'lying, untrustworthy, unwanted economic migrant'. On a second approach, a slur functions semantically to derogate, disparage, or cast aspersions on the object of the slur. And so the slur 'nigger', for instance, means or includes within its semantic content 'African American and despicable because of it'.⁵⁹ A third account maintains that slurs include within their semantic content the fact that the speaker possesses certain emotions, feelings, or attitudes, such as contempt, disdain or scorn, toward the objects of the slurs, and intends to use the slurs to express these states; to express the way the speaker looks on the objects of the slurs as worthless or beneath consideration.⁶⁰ So the slur 'Yao Ming is a Chink' means the following, 'The speaker believes that Yao Ming is Chinese, the speaker identifies being Chinese as part of what Yao Ming is, the speaker holds Yao Ming in contempt on account of being Chinese, and the speaker intends to express this contempt.'⁶¹

Second, most people, it seems to me, would also class as hate speech any forms of speech or expressive conduct that express or articulate ideas relating to the moral inferiority, lowness, non-humanity or alienness of members of historically oppressed or victimised groups or perhaps members of any groups identified by protected characteristics.⁶² This broad type of speech includes speech

⁵⁸ See, e.g., Lynne Tirrell, 'Racism, Sexism, and the Inferential Role Theory of Meaning', in C. Hendricks and K. Oliver (eds.) *Language and Liberation: Feminism, Philosophy, and Language*. Albany, NY: State University of New York Press, 1999). Cf. Robin Jeshion, 'Slurs and Stereotypes', *Analytic Philosophy* 54 (2013): 314–325.

⁵⁹ See, e.g., Christopher Hom, 'The Semantics of Racial Epithets', *Journal of Philosophy* 105 (2008): 416–440, p. 416.

⁶⁰ See, e.g., Jeshion, 'Slurs and Stereotypes'; Robin Jeshion, 'Slurs, Dehumanization, and the Expression of Contempt', in D. Sosa (ed.) *Bad Words* (Oxford: Oxford University Press, forthcoming). Cf. Hom, 'The Semantics of Racial Epithets'; Elisabeth Camp, 'Slurring Perspectives', *Analytic Philosophy* 54 (2013): 330–349.

⁶¹ See, e.g., Jeshion, 'Slurs and Stereotypes', pp. 240–243.

⁶² See, e.g., Matsuda, 'Public Response to Racist Speech', p. 2333; Molefi Kete Asante, 'Identifying Racist Language: Linguistic Acts and Signs', in M. Hecht (ed.) *Communicating prejudice* (Thousand Oaks, CA: Sage, 1998), p. 92; Parekh, 'Hate Speech', p. 214; Kevin W. Saunders, *Degradation: What the History of Obscenity Tells Us About Hate Speech* (New York, NY: New York University Press, 2011), pp. 100, 132; Musolf, 'Dehumanizing Metaphors in UK Immigrant Debates in Press and Online Media', p. 49.

that: ranks certain groups or classes of persons as morally inferior; compares them to non-human animals or to parasitic creatures; casts them as alien or not of (or belonging to) this world; draws connections between ascriptive characteristics and low distinctions of merit, dignity, status, and personhood. Consider this racist utterance, 'In my opinion blacks have depraved souls, they engage in acts of savagery, they are not part of the human species; so it is safe to say that whenever I come across them I know they don't merit my concern as moral beings, still less my admiration, trust, and friendship.' Or consider an image, sign, caricature or symbol (including an audio symbol) that depicts or represents certain races or ethnic groups as inferior or degraded – for example, the display of a caricature drawing of a monkey, the holding up of a banana (actual or plastic), or the making of monkey noises in front of, directed at or addressed to people who are perceived to be members of certain races or ethnic groups by people who take themselves to be members of a different race or ethnic group.⁶³

Now we also sometimes call such forms of expression 'hateful speech'. And it strikes me that when used in this way the term 'hateful speech' carries a second meaning, different to the one identified in 'The Myth of Hate'. This second meaning has little to do with the audience's reactions to speech, and even less with any feelings of hate or hatred that might be expressed by or motivate its usage. Instead, it has to do with the locutionary force of the speech, with what the speech *says* about members of certain groups. So, for example, to call racist speech 'hateful speech', in this other sense, is to comment on the ideational content of racist speech, as speech that conveys ideas relating to the moral inferiority, lowness, non-humanity, or alienness of certain races.⁶⁴ In other words, we sometimes use the terms 'hate speech' and 'hateful speech' to refer to speech that expresses certain ideas. But notice that here the word 'hate' does not carry its normal meaning (emotions, feelings, or attitudes of

⁶³ See, e.g., Matsuda, 'Public Response to Racist Speech', p. 2320; Delgado and Stefancic, *Understanding Words That Wound*, p. 141; Parekh, 'Hate Speech', p. 215.

⁶⁴ See, e.g., Matsuda, 'Public Response to Racist Speech', pp. 2366–2367; Anthony Lewis, *Freedom for the Thought that We Hate* (New York, NY: Basic Books, 2007), pp. 160–163; Corey Brettschneider, 'Value Democracy as the Basis for Viewpoint Neutrality: A Theory of Free Speech and Its Implications for the State Speech and Limited Public Forum Doctrines', *Northwestern University Law Review* 107 (2013): 603–645, p. 610.

extreme dislike or antipathy). Instead, it is much closer to the concept *contempt*.

Third, I believe that the average user of the term 'hate speech' would look upon group defamation, in the strict sense of making false damaging statements of fact which diminish the reputation or good standing of any members of groups or classes of persons identified by protected characteristics, as hate speech.⁶⁵ A classic example is the blood libel of Jews. 'All the Jews living in this town are complicit in the practice of kidnapping and murdering the children of Christians for the purposes of using their blood for Jewish holidays and rituals, like in the baking of matzos for Passover or for re-enacting the crucifixion.'

One common difficulty associated with group defamation as a type of hate speech is how to establish a connection between making false statements *about a group as a whole* and making false statements that are *of particular members of the group*. But assuming this can be resolved in particular instances, it does seem intuitive to call some Holocaust denial hate speech, for example, *qua* false statements of fact about Jews (as a group or about individual survivors or witnesses). By contrast, we would not normally label as 'hate speech' statements alleging that the 7/7 London bombings in 2005 were a hoax. This is presumably because the 'victims' of this sort of speech are people who were killed or injured in the bombings or the friends and family of people who were killed or injured and this is not a group who share some protected characteristic over and above the aforementioned. Similarly, if the 'target' of statements of the form, 'The so-called 7/7 London bombings were faked by the British authorities so as to demonise Muslims and sway public opinion in favour of their anti-terror policies' is the British authorities, then this seems to be more seditious libel than group libel.

It is a further question whether Holocaust denial can be plausibly labelled 'hate speech' even when it is not group defamation. What are the alternatives? Holocaust denial could also be understood as a negative stereotype about Jews. It could also be defined not by semantic content but by the sort of action involved, such as the act of inciting hatred against Jews or the act of affronting the dignity of (deceased) Jews. It might instead be classified as a *sui generis* form of hate speech. Indeed, some forms of Holocaust denial could poten-

⁶⁵ See, e.g., 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.

tially be considered something other than hate speech. There is a growing literature on how best to classify Holocaust denial.⁶⁶ But ultimately any plausible classification of Holocaust denial must depend on which particular statements one is talking about.⁶⁷ Maybe the following statement is hate speech in virtue of being either a defamatory statement of particular Jews or a negative stereotype about all Jews or Jews in general. ‘There were no gas chambers in German concentration camps during the period of the so-called Holocaust and the suppression of facts proving this to be the case is largely the result of a conspiracy perpetrated by certain powerful Jewish people and their wealthy institutions.’

Fourth, following on from the above example, I suspect that most ordinary language users would tend to treat as ‘hate speech’ negative stereotypes or generics about any members of groups or classes of persons identified by protected characteristics. Negative stereotypes typically depict people in a bad light. Yet they may not amount to false statements of fact of the sort that qualify as libel or defamation. This can be for a variety of reasons.⁶⁸ It could be because they express existentially quantified generalisations that are true because they are only saying that there is at least one member of the group in question who possess the bad attribute. Or because they express statements of statistical fact about the prevalence of certain attributes among the group in question and these statistical facts are true even if they also present an unbalanced, oversimplified, or misleading impressions of reality. Or because they express unquantified generalisations that are sufficiently vague, ambiguous, or metaphysical to make it impossible to determine whether they are true or false. Consider the following words published in *The Daily Express* during the mid-1990s by the TV personality and politician Robert Kilroy-Silk.

⁶⁶ See, e.g., Modood, ‘Muslims, Incitement to Hatred and the Law’, p. 145; Parekh, ‘Hate Speech’, p. 215; Raphael Cohen-Almagor, *The Scope of Tolerance: Studies on the Costs of Free Expression and Freedom of the Press* (London: Routledge, 2006), p. 12, Chap. 7; Raphael Cohen-Almagor, ‘Holocaust Denial is a Form of Hate Speech’, *Amsterdam Law Forum* 2 (2009): 33–42, p. 35; Kathleen Mahoney, ‘Hate Speech, Equality, and the State of Canadian Law’, *Wake Forest Law Review* 44 (2009): 321–351, p. 325; Frederick Schauer, ‘Social Epistemology, Holocaust Denial, and the Post-Millian Calculus’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012), pp. 142–143; Michel Rosenfeld, ‘Hate Speech in Constitutional Jurisprudence: A Comparative Analysis’, in M. Herz and P. Molnar (eds.) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press, 2012), p. 246; Brown, *Hate Speech Law*, pp. 96–98.

⁶⁷ Cf. Brown, *Hate Speech Law*, pp. 109–110.

⁶⁸ Cf. 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. 760–765.

Muslims everywhere behave with equal savagery. They behead criminals, stone to death female – only female – adulteresses, throw acid in the faces of women who refuse to wear the chador, mutilate the genitals of young girls and ritually abuse animals.⁶⁹

At face value these words might be read as universally quantified generalisations, describing all Muslims. If so, they are patently false. But faced with accusations of libel, Kilroy-Silk could retort that he simply meant to say that all of these things have been done by at least one Muslim in at least one part of the world where Muslims live. In which case the generics are actually existentially quantified generalisations, and they may turn out to be true. Or he means that these acts are, statistically speaking, more prevalent among Muslims than non-Muslims. Perhaps he could find a few statistics to back up these claims; or at least some of them. Or maybe he means to assert that these acts are characteristic properties of being a true Muslim as he conceives it. He might not be saying that all Muslims partake of these practices; merely that true Muslims do or that these practices define the essence of being a Muslim. But as metaphysical statements about what it means to be a true Muslim, it is hard to decisively judge them true or false. Or perhaps he simply means to say that these practices among some Muslims are notable or important practices, even if only a small proportion of actual Muslims engage in these practices and even if these practices do not define the essence of being a Muslim. If so, then it is almost impossible to prove that these are not notable or important practices, things of which notice ought to be taken. Finally, the meaning of his words may be irredeemably vague and ambiguous, and so their actual meaning cannot be definitively articulated so as to support judgements as to truth or falsity.

Fifth, I believe that people intuitively count as hate speech words or behaviour that, by intention or likelihood, incite, stir up or advocate hatred, discrimination or violence against members of groups or classes of persons identified by protected characteristics, or that actually threaten violence against such groups, or that simply justify or glorify discrimination or violence against such groups.⁷⁰

⁶⁹ Robert Kilroy-Silk, *Daily Express*, January 16, 1995.

⁷⁰ See, e.g., Frederick Schauer, 'Uncoupling Free Speech', *Columbia Law Review* 92 (1992): 1321–1357, p. 1349; Tsesis, *Destructive Messages*, p. 211, n. 1; Cohen-Almagor, *The Scope of Tolerance*, p. 11; Cohen-Almagor, 'Holocaust Denial is a Form of Hate Speech', p. 35; 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.

Consider the following Islamophobic speech, 'You think you can trust Muslims, think again, they are vile, backward, and dangerous people who deserve only our hatred, and when this country is finally united in its hatred of Muslims, they had better watch out!' Or, 'I will give anyone who kills a Muslim in our Christian land all the money I have.' Or, 'The day that those defenders of our values opened fire in that mosque was a brilliant day.'

Clearly some instances of hate speech can, and will, fall into more than one of the aforementioned action sub-types. For example, sometimes people use threatening words or behaviour in order to stir up hatred, but in some instances the same language could be further interpreted as actually threatening discrimination or violence and in other instances as simply justifying or glorifying discrimination or violence. Consider signs and symbols signifying historical and ongoing practices of discrimination or violence directed against members of minority groups. Commonly cited examples are burning crosses⁷¹ and swastikas.⁷² The pragmatics of such signs and symbols can be theorised in different ways depending on the context. It could be that when the main audience is likeminded or potentially like-minded people, they function to stir up hatred or to advocate discrimination or violence against those groups. But when the main audience is members of the targeted groups, these signs or symbols might serve to actually threaten discrimination or violence against those people. Or they could express solidarity with, or sympathy for, certain political regimes, institutions, and social groups that have perpetrated acts of discrimination or violence under the banner of the relevant signs or symbols, and thereby operate to justify or glorify those acts.

What is especially noticeable about this last broad type of hate speech is the fact that it is speech pragmatics rather than semantics which makes it distinctive: namely, the focus is more on what people are doing with the words or symbols they use as opposed to what they are saying or the ideational content of the words or symbols. In the next section I shall explore in more detail this general way of analysing speech, including all types of hate speech.

⁷¹ See, e.g., Matsuda, 'Public Response to Racist Speech', p. 2321; Lawrence 'Cross Burning and the Sound of Silence', pp. 787–788; Delgado and Stefancic, *Understanding Words That Wound*, p. 141; Heyman, *Free Speech and Human Dignity*, ch. 10; Waldron, 'Dignity and Defamation', p. 1600.

⁷² See, e.g., Matsuda, 'Public Response to Racist Speech', p. 2321; Cortese, *Opposing Hate Speech*, p. 1; Waldron, 'Dignity and Defamation', p. 1600.

VI. ORDINARY LANGUAGE ANALYSIS

It seems ordinary language users identify examples of hate speech not merely in virtue of what types of words people use or what types of things people *say* with words, but also in virtue of what people *do* with words. This implicitly taps into a specific sort of ordinary language analysis known as ‘speech act theory’. Drawing on J. L. Austin’s canonical distinction between three kinds of speech act,⁷³ it would be hard to deny, I think, that the term ‘hate speech’ can refer to certain ‘locutionary acts’, the act of saying something by using the words we do – for example, in calling someone a ‘Chink’ the speaker might be *saying* that he holds that person in contempt because he is Chinese or South East Asian. But it can also refer to ‘illocutionary acts’, the act of doing something by using the words we do – for example, through calling someone a ‘Chink’ the speaker might be performing the act of *disparaging* that person. And it can refer to certain ‘perlocutionary acts’, having an impact on another person by using the words we do – for example, by calling someone a ‘Chink’ the speaker might cause that person to *feel disparaged*.

To focus on the illocutionary acts, it seems that we can plausibly use the term ‘hate speech’ to describe a panoply of actions. Sometimes, perhaps often, we consider speech to be hate speech if it *perpetuates* or *facilitates* acts of subordination or oppression.⁷⁴ We might also consider speech to be hate speech if it actually *constitutes* acts of subordination or oppression.⁷⁵ Rae Langton identifies hate speech with various acts of subordination. Signs that read ‘Whites only’, for example, can serve to ‘*rank* blacks as having inferior worth’, ‘*legitimate* discriminatory behavior on the part of whites’, and

⁷³ See J. L. Austin, *How to Do Things With Words* (Oxford: Clarendon Press, 1962).

⁷⁴ See, e.g., Matsuda, ‘Public Response to Racist Speech’, pp. 2332, 2358, 2363.

⁷⁵ See, e.g., Andrew Altman, ‘Liberalism and Campus Hate Speech: A Philosophical Examination’, *Ethics* 103 (1993): 302–317, pp. 309–310; Rae Langton, ‘Speech Acts and Unspeakable Acts’, *Philosophy and Public Affairs* 22 (1993): 293–330, p. 303; 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; 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.

'deprive blacks of some important powers: for example, the power to go to certain areas'.⁷⁶ Racist propaganda espousing the inferiority of certain races 'incites and promotes racial discrimination, hatred and violence'.⁷⁷ Hate speech 'can be an illocution that persecutes and degrades'.⁷⁸ And when targeted at or addressed to a person or small group of persons hate speech can be an 'assault, insult, threat'.⁷⁹ In a similar vein, Vasu Reddy contends that homophobic hate speech is 'a discourse of power, dominance and control which is not merely a form of patriarchal oppression, but a kind of performative communication that produces a discourse about homosexuals in order to misrecognise them'.⁸⁰ On this kind of analysis, then, to engage in hate speech is not merely to speak, it is also to perform a 'speech act'⁸¹ or to participate in a 'social practice'.⁸² To further emphasise the performative nature of hate speech, some writers have coined the term 'assaultive speech' as an alternative to 'hate speech'.⁸³

What is much less certain, however, is which particular speech acts are and are not synonymous with hate speech. Some writers identify hate speech with *harming* or *inciting discrimination or violence*⁸⁴ or *inciting persecution*.⁸⁵ Brison characterises hate speech in terms of acts of *vilifying*, *intimidating*, *defaming*, and *inciting hatred*.⁸⁶ For her part, Katherine Gelber suggests that hate speech can be defined in terms of acts of *disempowering*, *marginalising*, and *silencing*.⁸⁷ These lists provide a good starting point but they are surely not exhaustive, nor intended to be so. Using speech act analysis we might come to the conclusion that the word 'nigger', for instance, can be used potentially to perform any or all of the acts on the

⁷⁶ Langton, 'Speech Acts and Unspeakable Acts', p. 303. See also Brown, *Hate Speech Law*, pp. 75–86.

⁷⁷ Langton et al., 'Language and Race', p. 758.

⁷⁸ *Ibid.*

⁷⁹ Langton, 'Beyond Belief', p. 77.

⁸⁰ Vasu Reddy, 'Perverts and Sodomites: Homophobia as Hate Speech in Africa', *Southern African Linguistics and Applied Language Studies* 20 (2002): 163–175, p. 164.

⁸¹ Langton et al., 'Language and Race', p. 758; Langton, 'Beyond Belief', p. 77.

⁸² Lawrence et al., 'Introduction', p. 1.

⁸³ *Ibid.* For a more critical examination of just how, and whether or not in fact, hate speech typically has the illocutionary force it is thought to have, see Butler, *Excitable Speech*.

⁸⁴ Schauer, 'Uncoupling Free Speech', p. 1349.

⁸⁵ Tsesis, *Destructive Messages*, p. 211, n. 1.

⁸⁶ Susan Brison, 'The Autonomy Defense of Free Speech', *Ethics* (1998) 108: 312–339, p. 313; Susan Brison, 'Hate Speech', in H. La Follette (ed.) *The International Encyclopedia of Ethics* (Oxford: Wiley-Blackwell, 2013), p. 2332.

⁸⁷ Gelber, 'Reconceptualizing Counter-speech in Hate Speech Policy', p. 213.

aforementioned lists plus a range of other acts including but not limited to *insulting, disparaging, degrading, humiliating, disheartening, harassing, persecuting, threatening, provoking, or inciting hatred, discrimination or violence*. A similarly long and varied list might be relevant to the act of burning crosses no doubt. But other forms of hate speech might have their own partly overlapping and partly unique lists. I have myself argued that some forms of hate speech can amount to various kinds of *misrecognising*.⁸⁸

It is also worth remembering at this stage that speech act theory incorporates not only a classification of speech acts but also an examination of the grammatical or syntactic properties of words along with critical reflection on whether there are any good reasons for their having these properties based on what we want to do with words. For example, the term 'hate speech' remains only a noun in English. It is not a verb; which is to say, few, if any, ordinary language users have taken to using the verb 'to hate speak against'. So the pressing question is whether there is any good reason why the term 'hate speech' itself is currently only used as a noun and not a verb given that to engage in hate speech can be to do things which can be described using the following verbs, 'to insult', 'to degrade', 'to defame', 'to negatively stereotype', 'to incite hatred, discrimination or violence'.

Perhaps some people will argue that there are already verb-equivalents of the term 'hate speech' in common usage, such as 'to hate on'. Yet this verb is used in relation to criticising anyone or saying derogatory things about anyone in a public way, and not necessarily about members of certain groups based on protected characteristics. Likewise, when people use the verbs 'to speak ill of' and 'to speak against' someone there is no limitation or restriction on the type of person being spoken ill of or spoken against. But the semantics of the term 'hate speech' would seem to imply that the verb 'to hate speak of' or 'to hate speak against' could only be used properly in relation to groups or classes of persons identified by protected characteristics. So in order to be accurate the relevant verb would need to be something like 'to hate speak against members of groups or classes of people identified by protected characteristics'. Perhaps, then, we should also keep in mind Austin's insistence that

⁸⁸ Brown, *Hate Speech Law*, pp. 166–174.

'ordinary language is *not* the last word: in principle, it can everywhere be supplemented and improved upon and superseded'.⁸⁹

That being said, it will require further analysis to explain what actions are and are not connected with this new verb.⁹⁰ And this, arguably, is where insights from both philosophy and sociolinguistics becomes important: it is a matter of documenting and analysing our existing uses of the term 'hate speech' to identify the sorts of acts that this term is and is not commonly used in connection with. Interestingly, Austin is also clear that philosophers and linguists engaged in the analysis of normative concepts should avail themselves of the law as another rich source of cases and uses of language.⁹¹ And it is not hard to find examples of legal documents, particularly in the field of human rights, which analyse hate speech in terms of some of the aforementioned speech acts.⁹² Consider the Council of Europe's characterisation of hate speech as forms of expression that 'spread', 'incite', 'promote', 'justify'.⁹³

VII. FAMILY RESEMBLANCES

In this section I want to defend the claim that the concept *hate speech* is what Wittgenstein called a 'family resemblances' concept.⁹⁴ According to Wittgenstein, the crucial point about family resemblances concepts like the concept *game*, for example, is that someone can be a competent user of the relevant term even if he or she cannot articulate a precise definition of that term (a set of necessary and jointly sufficient conditions for its correct usage) because the term does not admit of such a definition.⁹⁵ What really matters is that someone knows how to be a participant in the joint activity of using the term. Coming to have this knowledge involves being exposed to other users. As well as identifying the key features of the

⁸⁹ J. L. Austin, 'A Plea for Excuses: The Presidential Address', *Proceedings of the Aristotelian Society* 57 (1956–1957): 1–30, p. 11.

⁹⁰ *Ibid.*, pp. 5–6.

⁹¹ *Ibid.*, pp. 11–14.

⁹² Cf. Langton et al., 'Language and Race', p. 758.

⁹³ Recommendation No. R (97) 20, 30 October, 1997, Appendix. Available at [www.coe.int/t/t/dghl/standardsetting/hrpolicy/other_committees/dh-lgbt_docs/CM_Rec\(97\)20_en.pdf](http://www.coe.int/t/t/dghl/standardsetting/hrpolicy/other_committees/dh-lgbt_docs/CM_Rec(97)20_en.pdf).

⁹⁴ Wittgenstein, *Philosophical Investigations*, p. 36 [67].

⁹⁵ Note, however, the mere fact that the user lacks the knowledge does not prove that the necessary condition does not exist. See, e.g., Sorin Bangu, 'Later Wittgenstein on Essentialism, Family Resemblance and Philosophical Method', *Metaphysica* 6 (2005): 53–73.

joint activity, the present form of conceptual analysis also involves identifying sets of overlapping similarities that connect otherwise disparate examples of the given concept. In what follows I shall try to show that this method provides not only an illuminating way of understanding what Matsuda and other writers were doing when they provided examples of hate speech back in the late 1980s, but also a way of mapping and accounting for the internal diversity of hate speech.

Of course, it might be countered right off the bat that the project of analysing the ordinary concept *hate speech* requires at least some minimal characterisation of the term ‘hate speech’ which ordinary language users agree upon (explicitly or implicitly) even if their usage admits of a family of meanings. Otherwise the philosopher could not be sure that it is broadly one and the same concept that people have in mind when they use the term ‘hate speech’. Yet some writers doubt the existence of even this shared minimal characterisation.⁹⁶ However, it seems to me that this role could be, and probably is, played by paradigmatic examples or exemplars of hate speech – examples that any competent user of the term ‘hate speech’ is apt to classify as hate speech with little or no hesitation. What follows is a series of rational judgements, implicit or explicit, by ordinary language users that their use of the term ‘hate speech’ in any given case captures at least one important similarity with one or more paradigmatic examples. And this is surely part of what it might mean to say that hate speech is a family resemblances concept.

Based on everything I have said in this article and in ‘The Myth of Hate’, my working hypothesis is that even competent users of the term ‘hate speech’ lack knowledge of a set of necessary and jointly sufficient conditions for appropriately calling something ‘hate speech’. Now there are particular qualities, features, or descriptive properties that can be found in some or even in many examples of hate speech. Even so, these are not always present, and are not necessary conditions for applying the concept. For example, although many examples of hate speech involve verbal speech, many do not; many are about or against members of groups or classes of persons identified by race, yet there are also countless examples of non-racist hate speech; many instances of hate speech constitute

⁹⁶ See, e.g., Stanley Fish, ‘Going in Circles with Hate Speech’, *The New York Times*, November 12, 2012. Available at <http://opinionator.blogs.nytimes.com/2012/11/12/going-in-circles-with-hate-speech/>.

incitement to hatred and, then again, many do not. Nevertheless, what if we say that something is hate speech if and only if it is verbal or written speech or a symbol, gesture, or play, and is an insult or the expression of an idea relating to moral inferiority or defamation or a negative stereotype or the act of insulting or degrading or persecuting or threatening or inciting hatred, discrimination or violence against members of groups or classes or persons identified by their race or ethnicity or religion or nationality or gender or sexual orientation or disability? Surely this would mean that we have formulated a necessary condition: we are saying that all examples of hate speech share a complex disjunctive property.

However, there is something unsatisfying or inadequate about this suggestion. For one thing, we seem to want to know what single, *simple* quality or feature or descriptive property is common to all things that fall under the concept. The disjunctive property does not fit the bill of providing this *essence*.⁹⁷ Also, we might think that there is something *coherent* about the concept hate speech that makes it unlike any number of incoherent concepts we could invent by randomly combining properties together disjunctively. Moreover, in order for us to say that we know something that all examples of hate speech have in common even without having to look at all examples of hate speech, we would have to be able to set out the complete formulation of the disjunction in advance of looking at all examples of hate speech. But that seems like a tall order. It scarcely seems probable that we have at our finger tips a definitive list of all the forms of speech, all the kinds of symbols, all the types of speech act, and all the kinds of targeted characteristics that are associated with the concept *hate speech*. No doubt we could provide a set of extremely long lists, but for all we know there could be many things that are not included in the lists but which competent users of the term 'hate speech' would nevertheless treat as hate speech. Over time people apply the term 'hate speech' in unanticipated ways and the rate of new applications does not show signs of reaching zero any time soon. If competent users of the term 'hate speech' do not possess a priori knowledge of the complete disjunction, then this is devastating to the claim that they possess a priori knowledge of a set

⁹⁷ Cf. Wittgenstein, *Philosophical Investigations*, p. 37 [67].

of necessary and sufficient conditions for appropriately applying the term.

But do we really need an analytic definition of the term 'hate speech'? Does the concept *hate speech* become irredeemably artificial and incoherent without it? I would suggest not. One distinctive feature of family resemblances concepts is the fact that everything that falls under the concept shares at least one similar quality, feature, or descriptive property with at least one other thing that falls under the concept, even if there is no single quality, feature, or descriptive property that is common to all things that fall under the concept. This means that if we as competent users pause to look at examples that fall under a family resemblances concept, then, as Wittgenstein puts it, 'we see a complicated network of similarities overlapping and criss-crossing: similarities in the large and in the small'.⁹⁸ Thus, a competent user may know of many examples of hate speech and know of many similarities, but not know of one similarity shared by all the examples. This knowledge comes by learning how to be a participant in the complex activity of using the term 'hate speech'. It is a matter of acquiring what Oswald Hanfling has called 'participatory knowledge', which has to do with learning what to say and what not to say, and with participating alongside other speakers not only in using the term but also in giving and receiving sanctions for not using the term correctly.⁹⁹ This participatory knowledge prevents usage of the term from lapsing into incoherence.

But why, then, is it not the case that every form of speech can be subsumed under the concept *hate speech* given that every form of speech has at least one similarity with forms of hate speech, not least the similarity of speech? Wittgenstein approaches this problem by pointing to the way we explain things like games, or the family resemblances concept *game*, to other people. He writes:

⁹⁸ *Ibid.*, p. 36 [66]. The precise meaning of distinction that Wittgenstein is getting when he mentions 'similarities in the large and in the small' ('Ähnlichkeiten im Großen und Kleinen') is not clear-cut. 'Similarities in the large', for example, could be taken to mean: (1) qualities or features or descriptive properties that are common to a large number of things that fall under the concept; (2) that two or more things that fall under the concept share in common a large number of their qualities or features or descriptive properties; or (3) that two or more things that fall under the concept share in common a quality or feature or descriptive property that is important or encompassing or immediately noticeable when looked at.

⁹⁹ Oswald Hanfling, *Philosophy and Ordinary Language* (London: Routledge, 2000), p. 54.

How should we explain to someone what a game is? I imagine that we should describe *games* to him, and we might add: "This *and similar things* are called 'games'".¹⁰⁰

In other words, 'One gives examples and intends them to be taken in a particular way'.¹⁰¹ But Wittgenstein also makes it clear that he does *not* mean that someone 'is supposed to see in those examples that common feature which [the explainer] was unable to formulate',¹⁰² as though the examples are intended to be the vital clue to, or somehow stand-in as a shorthand for, the feature or features that are common to all games, or as though there is an essence to games after all. Instead, we are teaching the other person how to be a participant in the activity of using the term 'games'; which is to say, directing him or her to use it as we have done, namely, by using it with reference to the paradigmatic examples or exemplars we have highlighted and by making rational judgement calls as to what other things are similar and so can also be called 'games'.

This, I think, provides one plausible story about what it was that the original users of the term 'hate speech' were doing when they provided examples. Think about the sequence of events in which a language user or community of users focuses its attention on certain speech episodes and decides to dub them 'hate speech' (the moment at which the term is coined). In fact, it is likely that the term 'hate speech' first came into usage in April 1988 during an ADL-sponsored conference at Hofstra University organised by Monroe H. Freedman, then Dean of Hofstra Law School. At the conference numerous descriptions of speech events were put forward for discussion and some of the speakers called them 'hate messages',¹⁰³ 'the language of hatred', or simply 'hate speech'.¹⁰⁴ (Others used the older

¹⁰⁰ Wittgenstein, *Philosophical Investigations*, p. 37 [69].

¹⁰¹ *Ibid.*, p. 38 [71].

¹⁰² *Ibid.*

¹⁰³ Mari Matsuda, 'Outsider Jurisprudence: Toward a Victim's Analysis of Racial Hate Messages', in M. Freedman and E. Freedman (eds.) *Group Defamation and Freedom of Speech: The Relationship Between Language and Violence* (Westport, CT: Greenwood Press, 1995), pp. 87, 96.

¹⁰⁴ Kenneth Lasson, 'To Stimulate, Provoke, or Incite: Hate Speech and the First Amendment', in M. Freedman and E. Freedman (eds.) *Group Defamation and Freedom of Speech: The Relationship Between Language and Violence* (Westport, CT: Greenwood Press, 1995), p. 268.

term ‘group defamation’¹⁰⁵). Interestingly, it was an interview with Freedman that provided the basis for Tamar Jacoby’s 1988 article in *Newsweek*¹⁰⁶ which the Oxford English Dictionary cites as one of the first sources of the term ‘hate speech’.¹⁰⁷ For her part, Matsuda also used the terms ‘racist hate speech’ and ‘hate speech’ in her commentary at a roundtable discussion held in November 1988 at the State University of New York at Buffalo.¹⁰⁸ And in 1989 Matsuda published a revised version of her Hofstra conference paper in which she uses the term ‘hate speech’ repeatedly and systematically.¹⁰⁹ Matsuda’s article begins with a brief account of some examples of hate speech, including one event she experienced directly.¹¹⁰ And so what she goes on to explain about hate speech in the article might be paraphrased, in Wittgensteinian terms, as follows: ‘These *and similar things* are called ‘hate speech’ and I intend them to be part of the rules governing the social practice of using this term.’ Indeed, although all of Matsuda’s initial examples involve racist hate speech, she goes on to make it clear that in her judgement racist hate speech is not the only form of hate speech and that anti-gay and anti-lesbian speech are *similar things*.¹¹¹ Putting this another way, it might be possible to describe what Matsuda and others were doing as pioneering a complex norm-governed social practice of using the term ‘hate speech’ in some ways and for some examples but not for others.¹¹² Over time the repertoire of this practice has grown significantly, extending to types of examples not initially flagged up by Matsuda.

This raises a question as to whether users of the term ‘hate speech’ treat some elements of the repertoire, some members of the family resemblances concept *hate speech*, as more significant than others. Interestingly, in her work on the way human beings know or

¹⁰⁵ See Monroe H. Freedman and Eric M. Freedman (eds.) *Group Defamation and Freedom of Speech: The Relationship Between Language and Violence* (Westport, CT: Greenwood Press, 1995).

¹⁰⁶ Tamar Jacoby, ‘Time to Outlaw Racial Slurs? New Questions About Protecting Hateful Speech’, *Newsweek*, June 6 (1988): 48.

¹⁰⁷ Oxford English Dictionary, ‘Hate Speech’, July 2002. Available at www.oed.com.

¹⁰⁸ 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.

¹⁰⁹ Matsuda, ‘Public Response to Racist Speech’, pp. 2322–2323, 2326–2327, 2331–2334, 2347, 2363–2364, 2370–2378, 2380.

¹¹⁰ *Ibid.*, pp. 2320–2321.

¹¹¹ *Ibid.*, pp. 2331–2332.

¹¹² Cf. Robert Brandom, *Making It Explicit* (Cambridge, MA: Harvard University Press, 1994).

become aware of concepts at the level of mental cognition, Eleanor Rosch has sought to experimentally support the hypothesis that some members of family resemblances concepts, *prototypes*, are more central than others.¹¹³ This, put simply, means that people who understand the concept will tend to rank prototypes higher as good examples of the category and will recognise them as members of the concept more quickly, and will name them more frequently when asked to name a few exemplars. For example, when asked to give an example of the concept *furniture*, *chair* is more frequently cited than *lamp*. Even more interestingly, Beverley Fehr and James A. Russell have sought to demonstrate experimentally the idea that *love* is a concept that denotes a very large number of forms of love but that some forms of love, such as maternal love, friendship love, and romantic love, tend to be mentioned more often as forms of love, tend to be ranked higher as good examples of love, tend to be ranked more quickly as examples of love, and tend to be less frequently denied as being examples of love.¹¹⁴ Perhaps similar experiments could be conducted for the concept *hate speech* to see which, if any, forms of hate speech are prototypes, and whether these prototypes are or are not the very forms of hate speech that were first associated with the term 'hate speech' by Matsuda and others in the late 1980s. It would also be interesting to know whether any examples of the legal concept *hate speech* crop up as prototype or core examples of the ordinary concept *hate speech*.¹¹⁵ For example, it would be interesting to know whether ordinary language users would treat *incitement to hatred* as a prototype or core example of the ordinary concept *hate speech*, analogous to the way lawmakers and international organisations often seem to focus on this form of hate speech above other forms.

¹¹³ Eleanor Rosch, 'Cognitive Representation of Semantic Categories', *Journal of Experimental Psychology* 104 (1975): 192–233; Eleanor Rosch, *Wittgenstein and categorization research in cognitive psychology*, in M. Chapman and R. Dixon (eds.) *Meaning and the growth of understanding: Wittgenstein's significance for developmental Psychology* (Hillsdale, NJ: Erlbaum, 1987).

¹¹⁴ Beverley Fehr and James A. Russell, 'The Concept of Love Viewed From a Prototype Perspective', *Journal of Personality and Social Psychology* 60 (1991): 425–438.

¹¹⁵ In my book I identified ten clusters of laws that have been used to restrict, directly or indirectly, uses of hate speech. Brown, *Hate Speech Law*, Chap. 2. These furnish a legal concept *hate speech* which includes various more specific legal subconcepts *hate speech*, including *group defamation*, *negative stereotyping or stigmatisation*, *the use of insults or words expressing ideas of moral inferiority*, *incitement to hatred*, *speech that poses a threat to public order*, *denial or glorification of genocide*, *affront to dignity*, and *discriminatory harassment or intimidation*. Would ordinary language users recognise similar examples of hate speech, and would they recognise any of them as prototype or core examples of the ordinary concept *hate speech*?

But putting to one side this potentially interesting research agenda in the field of prototype theory and psycholinguistics, it is worth emphasising that a complicated patchwork of overlapping and criss-crossing similarities is likely to produce tremendous internal diversity. I have already touched upon some of the diversity of hate speech in this article as well as in 'The Myth of Hate'. We have seen that the ordinary concept *hate speech* encompasses: a family of different purposes including but not limited to highlighting forms of harmful speech, flagging up socially divisive forms of speech, identifying forms of speech that can undermine people's sense of equality, articulating civility norms, and labelling forms of speech that undermine democracy; a family of types of speech including but not limited to insults, slurs, and epithets, words that express or articulate ideas relating to the moral inferiority, group defamation, and negative stereotypes or generics; a family of types of speech act including but not limited to insulting, disparaging, degrading, humiliating, misrecognising, disheartening, harassing, persecuting, threatening, provoking, inciting hatred, discrimination or violence, and justifying or glorifying discrimination or violence; and a family of characteristics including but not limited to race, ethnicity, nationality, citizenship status, religion, sexual orientation, gender identity, and disability.

But this is only the tip of the iceberg. Other writers have observed that hate speech can take permanent or semi-permanent forms as well as transient or fleeting forms.¹¹⁶ It can exist in written words, but it can also exist in spoken words. And it can take the form of signs, symbols, gestures, images, pictures, plays, music, and video games.¹¹⁷ Hate speech can be comprised of well thought-out statements of fact or exaggerated hyperbole. It can be longwinded and convoluted tirades or short sharp monosyllabic insults or slurs. It can be formal lectures or off the cuff remarks. It can be sober or humorous, poetry or prose. In fact, it can take all manner of literary forms, anything from rhetorical questions to jokes, metaphors,

¹¹⁶ See, e.g., Waldron, 'Dignity and Defamation', pp. 1609–1610.

¹¹⁷ See, e.g., Delgado and Stefancic, *Understanding Words That Wound*, p. 11. Waldron has tentatively suggested that hate speech could be most problematic when it takes permanent or semi-permanent forms – published words, graffiti, posters, leaflets, badges on clothing, burning symbols and effigies. See Waldron, 'Dignity and Defamation', pp. 1603–1604, 1618. But then he immediately concedes that when hate speech in its transient forms – the spoken word, broadcast music – is repeated by enough people it can become a permanent or semi-permanent fixture of the sound environment. *Ibid.*, pp. 1623–1624.

myths, and proverbs.¹¹⁸ The style can be temperate or intemperate, emotive or calm, aggressive or passive, sincere or sarcastic.¹¹⁹ Hate speech is sometimes overt and unconcealed, but sometimes veiled and coded.¹²⁰ It can be done by an identified speaker or by an anonymous speaker.¹²¹ Hate speech can be used in private settings or public settings or in places that are both public and private. Hate speech can be performed by people in authority or by ordinary citizens, by professionals as much as by blue collar workers.¹²² Hate speech can be directed at or targeted toward, in the sense of being directly addressed to, an individual or individuals identified by protected characteristics, but it can also be about, and in a sense generally addressed to, an entire group or class of persons identified by such characteristics.¹²³ Hate speech is very often addressed to people who are members of the very groups that the relevant words are about or to which the words refer, but hate speech can also be addressed to people who are not members of the very groups in question such as in the case of incitement to hatred.¹²⁴ The targets of

¹¹⁸ See Matsuda, 'Public Response to Racist Speech', p. 2368; Laraine R. Fergensen, 'Group Defamation: From Language to Thought to Action', in M. Freedman and E. Freedman (eds.) *Group Defamation and Freedom of Speech: The Relationship Between Language and Violence* (Westport, CT: Greenwood Press, 1995), pp. 75–80; Asante, 'Identifying Racist Language: Linguistic Acts and Signs', pp. 92–93; Musolff, 'Dehumanizing Metaphors in UK Immigrant Debates in Press and Online Media', p. 49.

¹¹⁹ Parekh, 'Hate Speech', pp. 214–215.

¹²⁰ Delgado and Stefancic, *Understanding Words That Wound*, p. 11.

¹²¹ Cortese, *Opposing Hate Speech*, p. 1.

¹²² Delgado and Stefancic, *Understanding Words That Wound*, pp. 11–12.

¹²³ *Ibid.*, p. 11. Interestingly, several authors believe that hate speech is most regulatable when it is directly targeted at or addressed to a particular individual or individuals identified by protected characteristics as opposed to when it is about, and only indirectly or generally addressed to, an entire group or class of persons identified by such characteristics. 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; Kent 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 W. 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. 362–263; Caleb Yong, 'Does Freedom of Speech Include Hate Speech?', *Res Publica* 17 (2011): 385–403, pp. 394–396.

¹²⁴ In the case of a racist or ethnic epithet like 'chink', the word refers to Chinese people and when someone calls a Chinese person a chink to his or her face or in an online message, say, the hate speech is addressed to a person who is a member of the very group to which the word refers. The person is both the subject of hate speech and subjected to hate speech. By contrast, consider speech which is intended to stir up racial or ethnic hatred but which is actually addressed to an audience who do not belong to the racial or ethnic group against whom hatred is being stirred up. A union boss gives a speech to a group of white labourers in which he says, 'It is the duty of all good white citizens to unite against these Chinese coolies, who deserve only our hatred and contempt for working cheaper than we can suffer to work, cheaper than even a dog could work.'

hate speech either can be free to avoid it or they can be a captive audience.¹²⁵

Bearing this diversity in mind, it seems to me that when writers are criticised on the grounds that ‘their examples of hate speech do not self-evidently make up a ‘unitary’ phenomenon’,¹²⁶ we might respond on their behalf by asking, ‘Why must we assume that hate speech is a unitary phenomenon?’ Or, ‘Why must we suppose that the ordinary concept *hate speech* must be subject to the sort of precise and agreed definition as we might expect of the legal concept *hate speech*?’ No doubt similar amounts of internal variety are exemplified across the family of practices that ordinary language users are apt to call ‘games’.¹²⁷ But here we do not attack people for using the term ‘game’ for their inability to provide a precise definition.

I wish to make it clear at this point that the above family resemblances analysis is intended to apply first and foremost to the ordinary concept *hate speech*. The legal concept (or class of legal concepts) *hate speech* has itself inspired ordinary users to take up the term ‘hate speech’. But now that the genie is out of the bottle it has taken on a life of its own, and the ordinary concept *hate speech* may have outgrown the narrow, technical legal definition of ‘hate speech’ as incitement to hatred that is used in many legal jurisdictions. Therefore, I concur with Donna L. Lillian when she writes that the ‘critique’ that writers have failed to provide a ‘narrow definition of hate speech’ ‘is well taken if one’s primary concern is in dealing with legal and constitutional aspects of hate speech, but not necessarily if one’s primary concern is with linguistic, communicative, social, or cultural aspects of hate speech.’¹²⁸ The main purpose of the idea of family resemblances is to point out that in these other domains we do not need narrow definitions to know what we are talking about.

It is an open question whether the legal concept (or class of concepts) *hate speech* is also a family resemblances concept. Now it is certainly true that if we are talking about more specific, less general legal concepts that are to be used by courts, then we need to aspire towards having clear boundaries and a single unambiguous meaning.

¹²⁵ Delgado and Stefancic, ‘Four Observations About Hate Speech’, p. 362.

¹²⁶ John D. H. Downing, ‘“Hate Speech” and “First Amendment Absolutism” Discourses in the United States’, *Discourse and Society* 10 (1999): 175–189, p. 180.

¹²⁷ Wittgenstein, *Philosophical Investigations*, p. 36 [66].

¹²⁸ Donna L. Lillian, ‘A Thorn By Any Other Name: Sexist Discourse as Hate Speech’, *Discourse Society* 18 (2007): 719–740, p. 733.

The elements of specific civil or criminal offences must be couched in language that enables prosecutors, judges and juries to check off the elements of the offence against the circumstances of given legal cases. It is especially important for new and relatively artificial legal concepts that they can be given sufficient specificity to enable not only legal professionals but also the public at large to have a reasonable grasp of what they mean so that they can predict potential liability. For example, the legal concept *group libel* has a long history dating back at least as far as Roman defamation law which protected ruling groups,¹²⁹ and to the common law concept *sedition libel*.¹³⁰ But it has also evolved over time into various more specific, less general legal concepts such as *group libel (sensu stricto)*, *incitement to hatred*, *speech against protected groups or classes of persons that causes or is likely to cause a breach of the peace*.¹³¹ Our understanding of these concepts still reflects to some extent the older concept, but these are still relatively new pieces of legal conceptual equipment. The same goes for the legal concept *discriminatory harassment*. For relatively new pieces of legal conceptual equipment the challenge is to provide a definition that can be justiciable, that is, the subject of legal or quasi-legal proceedings in a court of law, tribunal, or hearing. So, if the main purpose of the idea of family resemblances is to point out that we do not need definitions to know what we are talking about, then arguably this purpose is out of place in the case of more specific, less general legal concepts like *incitement to hatred*. However, the same does not necessarily follow for the most inclusive or abstract legal concepts. These concepts enable us to talk about a broad range of more specific, less general concepts. And so we might think that it is appropriate for the general legal concept *hate speech* to operate as an umbrella or family resemblances concept, especially in instruments of intentional law which must somehow speak to a wide variety of domestic laws. To say that we do not need a narrow definition of the umbrella legal concept *hate speech*, however, is perfectly consistent with also saying that we *do* need narrow definitions of more specific, less general legal concepts such as *group*

¹²⁹ David Riesman, 'Democracy and Defamation: Control of Group Libel', *Columbia Law Review* 42 (1942): 727–780, p. 728.

¹³⁰ *Ibid.*, pp. 734–756.

¹³¹ See, e.g., Brown, *Hate Speech Law*, Chap. 2.

defamation, incitement to hatred, discriminatory harassment, and so forth.

VIII. IMPLICATIONS

In this two-part article I have argued that the concept *hate speech* is best conceived not as a compositional concept but as an equivocal idiom. I have also tried to excavate some of the multiple meanings of this concept by highlighting some of the jobs we expect it to perform, some of the folk platitudes that surround it, and some of the types of speech and speech act that exemplify it. Along the way I have also sought to explode the myth of hate. Finally, I have argued that the ordinary concept *hate speech*, and perhaps also the legal concept *hate speech*, is a family resemblances concept.

Why do these arguments matter? Taking them in reverse order, I believe that getting clearer about the nature of *hate speech* as a family resemblances concept may make us more alive to recognising as hate speech forms of speech that we might not have recognised as such had we been expecting a set of necessary and sufficient conditions. For example, if we had tried to build a narrow definition of hate speech based on the work of critical race theorists we may have ended up thinking that something is hate speech if and only if it is a racist insult. But the concept *hate speech* is useful precisely because it can be diversified to encompass an array of things that are both like and unlike racist insults, and which are like and unlike each other. The very fact that the concept *hate speech* denotes a complicated patchwork of overlapping and criss-crossing similarities renders it a far more flexible concept.

Indeed, it is worth remembering that a failure to recognise things as hate speech can have negative practical consequences. One is that when authorities or society as a whole do not even recognise certain forms of speech as hate speech it might serve to normalise, sanction, or authorise the relevant hate speakers to carry on doing what they are doing. Another is that it might lead authorities or society as a whole to neglect to take the necessary active steps to tackle hate speech, and this could mean that victims of unrecognised hate speech end up lacking parity of protection as compared to victims of recognised hate speech.

Of course, some people might be inclined to reject out of hand all hate speech regulations on the basis that there is no agreement about the meaning of the term 'hate speech'. How can we ban it if nobody is agreed about what it is? What I have tried to show, however, is that the term 'hate speech' has a family of meanings, and so any questions about whether or not and why we should ban hate speech must be addressed in the final analysis to particular members of the family. Indeed, we stand a much better chance of arriving at narrow definitions we can all share when we focus on particular members of the family. Once we understand this, then the disagreement cliché loses its power as a weapon against particular hate speech regulations.

In addition, I believe that if my arguments about the myth of hate are correct, then this has some important implications not merely for the way we think about the nature of hate speech, but also for what we say about how authorities should deal with hate speech, whether by instituting hate speech regulations and/or employing extralegal measures like education and counter-speech, or simply doing nothing at all. In the first instance, I suspect that exploding the myth of hate might under certain circumstances help to neuter at least two objections that have been levelled against hate speech regulations. First, some people are hostile to hate speech regulations because they think that it is not the business of the state to regulate people's emotions, feelings, or attitudes of hate or hatred, which are part of people's inner or private world. But if we dispel the misnomer that hate speech is essentially or necessarily connected with hate or hatred, then it may be easier to persuade such people that in fact regulating hate speech is not in the business of telling people what to feel or think. Second, some people are deeply ambivalent towards hate speech bans because they quite justifiably believe that the mere fact that an audience hates a form of speech is not in itself a legitimate reason to ban it. But once the myth of hate is exploded it becomes much harder to sustain the misleading idea that something qualifies as hate speech only if it is the sort of speech that we hate. And so hate speech laws do not seek to ban speech simply in virtue of the fact an audience hates it.

That being said, it could be pointed out that if what counts as hate speech is simply a matter of extension, that is, based on similarities

with exemplar cases, people will not be inclined to accept the regulation of hate speech unless they accept that the exemplar cases should be regulated. This seems fair. But here once again I think that exploding the myth of hate might be useful. I believe that even when Matsuda and other critical race theorists used the term 'hate speech' in reference to exemplar cases, they were not assuming or presupposing that all such speech is motivated by emotions, feelings, or attitudes of hate or hatred. For example, many of Matsuda's exemplar cases involve the use of racist slurs and insults against members of racial minorities, but Matsuda does not incorporate into her descriptions of these cases the crude assumption that the perpetrators were necessarily expressing or motivated by hate or hatred. What is more, both Matsuda and Delgado support the use of various kinds of hate speech laws in tackling the problem of hate speech – for example, the tort of intentional infliction of emotional distress, civil rights laws forbidding discriminatory harassment in the workplace, campus codes banning discriminatory harassment in universities¹³² – and, what is more, actively propose the 'creation of a new crime of racist speech'¹³³ or a new 'tort for racial insult'.¹³⁴ Yet at the same time none of these laws specify being motivated by hate or hatred among the basic elements of the relevant offences. Once critics of hate speech laws reflect on these facts surely it becomes much harder to object to hate speech laws on the grounds that they are about regulating people's emotions, feelings, or attitudes of hate or hatred.

Nonetheless, what about incitement to hatred laws? Are not such laws explicitly about controlling people's innermost emotions, feelings, or attitudes by force of law? Indeed, even Matsuda suggests that the main point and purpose of incitement to hatred law is 'avoiding the spread of hatred'.¹³⁵ This is certainly one way to look at these laws, but it is by no means the only way. Another way starts from the basic fact that these laws do *not* ban hate or hatred as such; what they ban is the act of stirring up these potentially pernicious states of mind. (Interestingly, some people object to incitement to hatred laws precisely because these laws ban the stirring up of hatred against

¹³² See, e.g., Matsuda, 'Public Response to Racist Speech', p. 2336, n. 83; 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. 133.

¹³³ Matsuda, 'Public Response to Racist Speech', p. 2360, n. 207.

¹³⁴ Delgado, 'Words That Wound', pp. 179–180.

¹³⁵ Matsuda, 'Public Response to Racist Speech', p. 2347.

groups or classes of persons identified by protected characteristics when there are no parallel laws banning merely hating people because of their possession or perceived possession of protected characteristics.¹³⁶) Thus, it might be argued that the real function of incitement to hatred laws is to protect people's autonomy. In short, the aim is not for the state to control people's emotions, feelings, or sentiments but instead for the state to intervene to prevent some citizens from illegitimately seizing control of other citizens' emotions, feelings, or sentiments.¹³⁷ This comes out in the English case *R. v. el-Faisal*¹³⁸ involving the prosecution of a Muslim cleric for *inter alia* using threatening, abusive or insulting words or behaviour with the intention of stirring up racial hatred against Jews. In his sentencing remarks Judge Beaumont touched upon el-Faisal's position as a cleric, with the power to shape the minds of 'young and impressionable' people within the local Muslim community.¹³⁹ To clarify, my main point here is not that incitement to hatred has nothing to do with hatred. Nor that hate speech properly called is never connected to hatred. Both of these claims are plainly false. Instead, my point is that getting clearer about the precise nature of the contingent connection between the concept *hate speech* and the concept *emotions, feelings, or attitudes of hate or hatred* can shed new light even on forms of hate speech, such as incitement to hatred, which *are* closely connected with hatred, and in ways that might undercut certain objections to laws banning hate speech.

Another of my main aims has been to question the assumption that the best way to explain the meaning of the term 'hate speech' is compositionally, drawing on a combination of literal and non-literal meanings of the term 'hate'. But even though I think the best way to understand the concept hate speech is non-compositionally, I do not think that we should abandon the term 'hate speech' simply because the word 'hate' is misleading. For one thing, it would be wrong, I think, to say that the word 'hate' is grossly misleading. The term 'hate speech' is used to refer to a wide variety of speech, and

¹³⁶ See, e.g., David Nash and Chara Bakalis, 'Incitement to Religious Hatred and the 'Symbolic': How will the Racial and Religious Hatred Act 2006 work?', *Liverpool Law Review* 28 (2007): 349–375, p. 367.

¹³⁷ See, e.g., Brown, *Hate Speech Law*, pp. 58–66.

¹³⁸ No. T20027343 (Central Criminal Court, March 7, 2003).

¹³⁹ Transcript obtained from Smith Bernal Reporting Ltd.

sometimes, as I have said, this speech is closely connected with emotions, feelings, or attitudes of hate or hatred. So the word 'hate' gets us thinking in not wholly inappropriate ways about hate speech in numerous instances. If hate speech never involved hate or hatred, then the case for abandoning the term would be undeniable. So the word 'hate' is not grossly misleading. But it is still misleading. And it is misleading because other instances of hate speech do not involve and have no connection with hate or hatred. Many instances have instead a close connection with contempt, disdain, scorn, or condescension. Some instances are related to feelings or sensations of disgust or revulsion. And yet more instances are bound up with a desire or emotional need on the part of the speaker to be accepted as part of an in-group or even a basic wish to be noticed or the subject of public attention. Then again, some instances might have nothing to do with any of these things.

Now it might be further objected at this stage that my own arguments support a different conclusion to the one I have presented. Rather than try to rehabilitate our understanding of the semantics of the term 'hate speech', surely I should be agitating for societies to abandon this term entirely. Would not it be simpler and more responsible to tell people to stop using the term 'hate speech' rather than seeking to educate them about the ways in which the word 'hate' can be misleading?

I think not, and for three main reasons. First, even though the term 'hate speech' invites the myth of hate it is not as though we have plenty of other words to choose from which are not misleading in one way or another. If we used a word such as 'contempt' to create 'contempt speech', it might give the misleading impression that hate speech is essentially and necessarily connected with the emotion, sentiment, or attitude of looking down on something, regarding something as unworthy of one's consideration or respect. This might well be involved in many instances of hate speech, but it will not be involved in all instances. Similarly, if we use the word 'prejudiced' to create 'prejudiced speech', it might give other misleading impressions. And so the term 'hate speech' can be a serviceable substitute for a less misleading but equally elegant descriptor for the relevant family resemblances concept that we have yet to think of. Of course, in the case of the legal concept *hate speech*

the term 'hate speech' replaced the term 'group libel'. So maybe some other, better term will come along and displace 'hate speech'. In that sense debunking the term 'hate speech' might turn out to be a transitional strategy.

Second, if people (like me) who acknowledge the problem of hate speech and believe that at least some hate speech laws can be warranted all things considered (even if this means some compromise of principles) begin to argue for societies to drop all usage of the term 'hate speech' just at the moment the concept is gaining popular understanding, then it could be a case of 'be careful of what you wish for'. If hate speech is a problem that we have an obligation to do something about, then we need some term or other to refer to it, to provide a focal point for campaigners and pressure groups, and to give legislators something to wrap their heads around. If we lose the term 'hate speech', it might give the false impression that there is no problem, and the impetus to do something about it could be lost. So we carry on with the term 'hate speech' whilst at the same time being mindful of its limitations.

Third, I would point out that the term 'hate speech' can continue to play various important roles even once the myth of hate has been exploded. Even if we come to realise that hate speech does not necessarily have to do with emotions, feelings, or attitudes of hate or hatred, surely we can still use the term 'hate speech' to identify forms of speech that disproportionately harm already disadvantaged or victimised members of society, or that are destructive of social cohesion, or that run contrary to fundamental democratic values. Likewise, even if we have rejected the myth of hate we can nevertheless use the term 'hate speech' to express disapproval of certain forms of speech and to flag them up as regulatable.

Finally, I want to draw out some implications of my claim that the legal concept *hate speech* and the ordinary concept *hate speech* have come to something of a parting of the ways. It seems to me that if a society thinks that the legal concept *hate speech* is the only concept of hate speech, then it may also be inclined to stick with a narrowly restricted definition of hate speech as incitement to hatred, and may even fall into thinking that the only choice for authorities in responding to hate speech is either the criminal law or doing nothing. However, if, as I believe it should, a society comes to realise

that the ordinary concept *hate speech* denotes a heterogeneous collection of phenomena held together by family resemblances rather than a single essential feature, then there is no earthly reason for that society not to start to look for a plurality of ways of responding to hate speech, and to embrace the idea that the criminal law is not the only way of combating hate speech.

I offer two illustrations of this implication. First, the European Commission Against Racism and Intolerance (ECRI) has recently issued its General Policy Recommendation No. 15 on Combating Hate Speech.¹⁴⁰ Part of the purpose of this recommendation is to qualify and supplement recommendations set out in General Policy Recommendation No. 7 on National Legislation to Combat Racism and Racial Discrimination from 2002.¹⁴¹ The latter recommends to the governments of the 47 member States of the Council of Europe that they should enact, implement, and enforce criminal laws that penalise various types of hate speech.¹⁴² For its part, General Policy Recommendation No. 15 highlights some of the limitations of using the criminal law in combating hate speech – such as the danger that criminal laws can be used against the very minorities they are designed protect – and recommends some *additional* steps that governments should take to combat hate speech besides criminal law. These steps include: develop educational programmes for children, young persons, public officials and the general public that discourage the use of hate speech; engage in governmental counter-speech against hate speech; promote and encourage civil society and victims to engage in counter-speech against hate speech; provide assistance to victims on how to cope with the psychological effects of hate speech and give information to victims about their legal rights; encourage and promote forms of self-regulation of hate speech by public and private institutions; ensure effective use is made of the powers to combat hate speech by media and Internet regulators; withdraw material support by public bodies for political parties and other organisations that use hate speech. Given this, it is logical that

¹⁴⁰ ECRI, General Policy Recommendation No. 15 On Combating Hate Speech, December 8, 2015, Strasbourg. Available at www.coe.int/t/dghl/monitoring/ecri/activities/GPR/EN/Recommendation_N15/REC-15-2016-015-ENG.pdf.

¹⁴¹ ECRI, General Policy Recommendation No. 7 on National Legislation to Combat Racism and Racial Discrimination, December 13, 2002, Strasbourg. Available at www.coe.int/t/dghl/monitoring/ecri/activities/GPR/EN/Recommendation_N7/ecri03-8%20recommendation%20nr%207.pdf.

¹⁴² *Ibid.*, s. 18.

GPR No. 15 adopts a relatively broad definition of hate speech, one that encompasses much more than the common legal definition of hate speech as incitement to hatred.

[H]ate speech is to be understood for the purpose of the present General Policy Recommendation as the advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of ‘race’, colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status; Recognising that hate speech may take the form of the public denial, trivialisation, justification or condonation of crimes of genocide, crimes against humanity or war crimes which have been found by courts to have occurred, and of the glorification of persons convicted for having committed such crimes [...]¹⁴³

This definition not merely reflects the plural nature of hate speech as a family resemblances concept but also renders more coherent ECRI’s own recommendation for a multidimensional approach to tackling the problem of hate speech. (It also enables ECRI to speak to a variety of member states with diverse experiences of, and diverse legal and extra-legal approaches to tackling, the problem of hate speech.)

Second, Internet companies are increasingly taking interest in the need to limit the use of hate speech on their platforms and websites. It is a matter of debate whether this is ultimately motivated by commercial interest as opposed to loftier concerns to protect people from harm. But either way, what is clear is that when it comes to the definition of the term ‘hate speech’ some Internet companies are adopting definitions that go beyond narrow legalistic definitions. Compare YouTube and Facebook. YouTube is a video sharing Internet platform that enables users to post videos to be viewed and commented on by other users. It is not primarily used as a social networking website or messaging service. It is more akin to a public square for showing videos to anyone and everyone. It is because of this, I think, that YouTube’s ‘community guidelines’ characterise hate speech in relatively narrow terms, akin to legal definitions of hate speech. ‘Hate speech refers to content that promotes violence or hatred against individuals or groups based on certain attributes’.¹⁴⁴ By

¹⁴³ ECRI, General Policy Recommendation No. 15 On Combating Hate Speech, at p. 3.

¹⁴⁴ Available at <https://support.google.com/youtube/answer/2801939>.

contrast, Facebook is a social networking website that permits users to post content on personal and group profile pages and to send messages to other users in reasonably targeted ways. As such, Facebook's 'community standards' recognise and disallow two main forms of hate speech. On the one hand, the standards address forms of hate abuse that are often directed at particular users. 'Facebook removes hate speech, which includes content that directly attacks people based on their: race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity, or serious disabilities or diseases.'¹⁴⁵ On the other hand, there are forms of more generalised incitement to hatred that are addressed to a wider audience, whether an audience of like-minded people or would-be followers. 'Organisations and people dedicated to promoting hatred against these protected groups are not allowed a presence on Facebook'.¹⁴⁶ Indeed, as a direct response to high-profile complaints made by pressure groups about its failure to enforce its rules against hate speech, particularly hate speech of a personalised nature directly targeted at female users, in May 2013 Facebook posted a statement that acknowledged these complaints, set out a series of new steps for tackling the problem, and clarified its position on the meaning of the term 'hate speech'.

While there is no universally accepted definition of hate speech, as a platform we define the term to mean direct and serious attacks on any protected category of people based on their race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or disease. We work hard to remove hate speech quickly, however there are instances of offensive content, including distasteful humor, that are not hate speech according to our definition. In these cases, we work to apply fair, thoughtful, and scalable policies. This approach allows us to continue defending the principles of freedom of self-expression on which Facebook is founded. We've also found that posting insensitive or cruel content often results in many more people denouncing it than supporting it on Facebook. That being said, we realize that our defense of freedom of expression should never be interpreted as license to bully, harass, abuse or threaten violence. We are committed to working to ensure that this does not happen within the Facebook community. We believe that the steps outlined below will help us achieve this goal.¹⁴⁷

What this shows, I think, is that Internet companies are adopting definitions of the term 'hate speech' which are, within the broad

¹⁴⁵ Available at www.facebook.com/communitystandards#.

¹⁴⁶ *Ibid.*

¹⁴⁷ Available at www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054.

family of meanings associated with the concept, tailored to fit the particular forms of speech and communication that occur on given websites and platforms and that are generally recognised as hate speech by the relevant users and by high-profile campaigners and pressure groups. In the case of Facebook's clarification of policy in May 2013, for example, the company was responding to users and campaigners who were essentially saying, 'Misogynist abuse is sufficiently *similar* to other things you call 'hate speech' that you really must take it more seriously and work harder to remove it as you do other hate speech.'

ACKNOWLEDGEMENTS

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 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.