THE POLITICS OF HATE SPEECH LAWS

Alexander Brown and Adriana Sinclair

ROUTLEDGE

The Politics of Hate Speech Laws

This book examines the complex relationship between politics and hate speech laws, domestic and international. How do political contexts shape understandings of what hate speech is and how to deal with it? Why do particular states enact hate speech laws and then apply, extend or reform them in the ways they do? What part does hate speech play in international affairs? Why do some but not all states negotiate, agree and ratify international hate speech frameworks or instruments? What are some of the best and worst political arguments for and against hate speech laws? Do political figures have special moral duties to refrain from hate speech? Should the use of hate speech by political figures be protected by parliamentary privilege? Should this sort of hyperpolitical hate speech be subject to the laws of the land, civil and criminal? Or should it instead be handled by parliamentary codes of conduct and procedures or even by political parties themselves? What should the codes of conduct look like?

Brown and Sinclair answer these important and overlooked questions on the politics of hate speech laws, providing a substantial body of new evidence, insights, arguments, theories and practical recommendations. The primary focus is on the UK and the US but several other country contexts are also explored and compared in detail, including: Nigeria, Kenya, South Africa, India, China, Japan, Turkey, Germany, Hungary and Italy. Methodologically, the two authors draw on approaches and concepts from a range of academic disciplines, including: law and legal theory, political theory, applied ethics, political science and sociology, international relations theory and international law.

Alexander Brown is Reader in Political and Legal Theory at the University of East Anglia (UEA), UK. He is the author of *A Theory of Legitimate Expectations for Public Administration* (Oxford University Press, 2017), *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). He has published articles on the concept of hate speech, the 'Who?' question in the hate speech debate, the use of civil torts for hate speech, the nature of online hate speech and its captive audiences and precautionary approaches to hate speech regulation. In 2018 he was a Visiting Scholar at the University of Queensland.

Adriana Sinclair is Lecturer in International Relations at the University of East Anglia (UEA), UK. She is the author of *International Relations Theory and International Law: A Critical Approach* (Cambridge University Press, 2011). She has published articles on interdisciplinary approaches to international relations and international law, constructivism and the agency of international law. She is currently working a new comprehensive theory of international relations and international law. In 2012 she was an AHRC/BBC Radio 3 New Generation Thinker, and in 2018 she was a Visiting Scholar at the University of Queensland.

The Politics of Hate Speech Laws

Alexander Brown and Adriana Sinclair

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2020
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Brown, Alexander, 1974 December 27– author. | Sinclair, Adriana, 1977– author.
Title: The politics of hate speech law / Alexander Brown and Adriana Sinclair.
Description: Abingdon, Oxon ; New York, NY : Routledge, 2019.
Identifiers: LCCN 2019016505 | ISBN 9781472439147 (hardback) | ISBN 9781315553917 (ebook)
Subjects: LCSH: Hate speech—Law and legislation. | Hate speech—Law and legislation—Great Britain.
Classification: LCC K5210 .B767 2019 | DDC 345/.025—dc23
LC record available at https://lccn.loc.gov/2019016505

ISBN: 978-1-4724-3914-7 (hbk)
ISBN: 978-1-315-55391-7 (ebk)

Typeset in Galliard
by Apex CoVantage, LLC

For Eli, Delphine and Milo

Contents

Acknowledgements

This book was a team effort in a larger sense than just the collaboration of its two authors. As we juggled our time on the book we also juggled family time with our three children – Eli, Delphine and Milo – whose exuberance never fails to surprise, delight and distract (in a good way). We love you kids.

Chapter 1 reproduces two paragraphs from Alexander Brown, 'How in the Future Online Hate Content Might Be Both Created and Countered by Algorithms', *Huffington Post UK*, October 4, 2018.

Chapter 2 would not have been possible without suggestions of important historic events, political figures, legal cases and much more besides from Joanna Botha, Arife Kose, Ra Mason and Akin Oyawale, as well as from scores of UEA postgraduate students from around the globe, especially from Europe, Africa and Asia, who have over a number of years in Alexander Brown's Free Speech module contributed to mutual learning about the nature and importance of context in debates on hate speech.

Chapter 3 is a revised version of Alexander Brown, 'The Politics Behind the Introduction of Stirring Up Religious Hatred Offences in England and Wales', *Politics, Religion, and Ideology* 18 (2017): 42–72. We thank the journal's publisher for permission to reproduce this material. We would also like to thank Sally Broughton Micova, Charles Clarke, Michael Frazer, Stephen Greasley and Chris Hanretty for useful feedback on an early draft. Moreover, we are very grateful to the journal's two anonymous reviewers for their insights and suggestions.

Chapter 4 was presented to the UEA international politics and security research group, with thanks to Simon Curtis, Suzanne Doyle, Lee Jarvis, Lee Marsden, Ra Mason and Reilly Willis.

Chapter 5 is based on a presentation given at the 2012 IPSA World Congress in Madrid. We are especially grateful to input from Katharine Gelber and Wayne Sumner. We also thank Simon Thompson for comments on a draft and Alexandria Innes for reminding us of the protection of political beliefs in refugee law.

Chapter 6 benefited from literature suggestions courtesy of Richard Delgado.

Chapter 7 was presented at the International Network for Hate Studies Biennial Conference in Oshowa, Ontario, May 2018; the School of Political Science and International Studies, the University of Queensland, Research Seminar Series in Brisbane, August 2018; and as the Keynote Address at the Florida Golf Coast University conference on free speech, March 2019. We'd like to thank in particular Mohamad Al-Hakim, Katharine Gelber, Steven Heyman, Michael Frazer and Louise Sullivan, for input on several of the examples and arguments discussed.

Chapter 8 includes some suggestions from Alexander Tsesis concerning US legal cases dealing with political speech.

Alexander Brown
Adriana Sinclair
Norwich, March 10, 2019

1 Introduction

1.1 Introduction

Hate speech is a linguistic, social, cultural, technological and legal issue. But at the same time it is also a highly political and politicised issue. In politics one side might accuse the other of cynically exploiting hate speech for political gain, whilst the other side may respond by accusing the first of misusing hate speech laws to silence political viewpoints it simply does not like. In the current climate it can be hard to see any way out of the multilevel disputes that surround hate speech. What is it? What sort of problem, if any, does it represent? What, if anything, should be done about it?

As an example of this politicisation, one might be forgiven for assuming that all sides would agree that it is morally unacceptable to say things that might be offensive to racial or religious groups, they simply disagree on what to do about it. But that is simply not the case. In 2017 a Cato Institute/YouGov public opinion survey in the US found that among Democrats only 14 per cent thought that it is morally acceptable to say things that might be offensive to racial or religious groups, whereas among Republicans 24 per cent thought it was morally acceptable (Cato Institute/YouGov 2017, 12). Perhaps what motivates these particular Republicans is a wider concern that it has become all too easy to regard language as offensive. This view is perhaps encapsulated in the words of Republican Senator Steve King: 'White nationalist, white supremacist, Western civilization – how did that language become offensive?' (Stolberg 2018).

Of course, many people would argue that hate speech is not the same thing as merely offensive language. Hate speech is more extreme, it can be threatening in nature, for example. But once again not everyone agrees there is something morally unacceptable about using threatening language towards racial or religious groups. To give a UK example, the English Defence League (EDL) is a far-right organisation known for its anti-Islam and Islamophobic ideology and for its street demonstrations. In 2011, a member of the EDL, Guramit Singh, gave an impassioned speech to EDL supporters on the streets of Peterborough which ended with the following thinly veiled threat (in the context of the violence, rioting and street disturbances that have often accompanied EDL protests). 'I've got one thing to say to the members of the Islamic community who are Islamists [. . .] English Defence League 2011 we're coming to a street near you, we ain't even fucking started yet!'[1] Singh was later arrested on suspicion of committing a religiously aggravated public order offence.[2] EDL's leader, Tommy Robinson, is reported to have said this about the arrest. 'The EDL are fully behind him and we don't think there was anything wrong with his speech' (Reville 2010). At the time of writing Tommy Robinson is an advisor to the leader of the UK Independence Party (UKIP), a mainstream party in UK politics.

Of course, it might be tempting at this point simply to dismiss these sorts of views as fringe or extreme. But that assessment would be too quick. There are many mainstream liberal thinkers who argue that whether or not one finds it morally acceptable to say things that might be offensive to racial or religious groups, making accusations of 'hate speech' against people is itself morally problematic. Just as some critics of hate speakers accuse them of spite, vindictiveness and hatred toward racial or religious groups, so critics of critics of hate speakers accuse them of spite, vindictiveness and hatred toward so-called hate speakers. The American liberal philosopher Ronald Dworkin once referred to proponents of hate speech laws as 'fanatical moralists with their own brand of hate' (Dworkin 2009, ix).

The politicisation of the problem of hate speech is perhaps at its sharpest when politicians and political parties trade accusations of hate speech. To take one example, in the US midterm elections in November 2018 Rashida Tlaib and Ilhan Omar became the first two female Muslim congresswomen. They are both Democrats. Not long after being in office Omar was accused by some people, including Republicans and Democrats, of anti-Semitism. Omar had made a series of tweets (some old, some new, and some subsequently deleted) in which she claimed that 'Israel has hypnotized the world' and alleging that support for Israel among many congresspersons was due to the money ('Benjamins') donated by wealthy Jewish individuals and organisations (Crow 2019). Omar's accusers argued that her tweets perpetuated familiar anti-Semitic conspiracy theories. Yet in response Omar protested that accusations of anti-Semitism against her were themselves fuelled by or constituted Islamophobic negative stereotypes, aimed at silencing her.

> What I'm fearful of is that, because Rashida and I are Muslim, that a lot of our Jewish colleagues, a lot of our constituents, a lot of our allies, go to thinking that everything we say about Israel to be an anti-Semitic because we are Muslim [sic]. And so to me it is something that becomes designed to end the debate, because you get in this space of, yes, right like I know what intolerance looks like and I'm sensitive when someone says, 'The words you used Ilhan are resemblance of intolerance' [sic]. And I am cautious of that and I feel pained by that. But it's almost as if, every single time we say something regardless of what it is we say that is supposed to be about foreign policy or engagement, our advocacy about ending oppression, or the freeing of every human life and wanting dignity, we get to be labelled something and that ends the discussion, because we end up defending that and nobody ever gets to have the broader debate of what is happening with Palestine. I want to talk about, I want to talk about the political influence in this country that says it is okay for people to push for allegiance to a foreign country.[3]

Omar's oblique reference to divided allegiance, however, attracted yet further accusations of anti-Semitism, based on the idea that she was eluding to longstanding tropes about Jewish Americans having conflicted allegiances. Reflecting the growing controversy about Omar but also a much wider backlash against hatred in American politics, in March 2019 the House of Representatives took the unusual step of passing (by a margin of 407 to 23 votes, with Omar among the 407 Representatives voting in favour) a wide-ranging Resolution condemning various kinds of racist, anti-Semitic and Islamophobic acts and statements 'as hateful expressions of intolerance that are contradictory to the values that define the people of the United States'.[4]

At the same time, however, not all of Omar's critics – and not all critics of the Democrats' handling of the controversy – were perceived as having the necessary moral credibility or

ethical standing to call out the problem of hatred in American politics. For example, during the controversy President Trump took to Twitter to declare the following.

> It is shameful that House Democrats won't take a stronger stand against Anti-Semitism in their conference. Anti-Semitism has fueled atrocities throughout history and it's inconceivable they will not act to condemn it![5]

Yet many critics of Trump decried this and other similar interventions as an act (tweet) of breathtaking hypocrisy, from a president who has consistently exploited hateful rhetoric against various minority groups and who has allegedly defended white supremacists (Haaretz 2019).

Interestingly, one of the strategies used by Trump to deflect criticism away from his own hate speech has been to denounce 'the Mainstream Media' and 'the Fake News' as 'anti-Trump haters'.[6] His aim is to suck the force from criticisms that he is a hate speaker by tarring with the same brush some of the people who make that criticism, meanwhile hoping that the audience will accept the false equivalence between genuine hate speakers and 'anti-Trump haters'. It is in this vein that in July 2019 Trump Tweeted – in response to accusations made against him by Omar and many other political figures of racism and xenophobia for having earlier Tweeted that members of the so-called squad should 'go back' to where they came from – that '[t]he Democrat Congresswomen', including Omar, 'have been spewing some of the most vile, hateful, and disgusting things ever said by a politician in the House or Senate'. For its part, the House of Representatives responded to Trump's 'go back' Tweet by passing a further Resolution (by a narrow margin of 240 to 187 votes, and without the support of the Republican Party) that 'condemns President Donald Trump's racist comments'.

In addition to heated political disputes about what counts as hate speech, it can also seem as though there is no escaping an extreme polarisation of views on what to do about the problem. On the one side, there are many liberals, progressives and people on the political left who think that hate speech is a genuine problem for individuals and communities and that hate speech laws are necessary – whether that be in the form of criminal laws, civil laws, anti-discrimination laws or human rights laws. On the other side, there are some liberals as well as social conservatives and people on the political right who believe that hate speech is simply a concocted idea, a strain of political correctness, and that hate speech laws are simply political weapons being used to silence critics in entirely illegitimate ways. Consequently, it can strike people as naïve to imagine there could be any escape from the binary choice between thinking that legal measures can, and should, be among the battery of measures that authorities use against hate speech and thinking that whatever authorities do to combat hate speech they should never make it unlawful, not least because this puts authorities on a slippery slope to massive censorship (Strossen 1990, 2018; Chemerinsky 2003; Hare 2006; Heinze 2009; Baker 2009) (see Chapter 5).

Furthermore, even if everyone accepted the optimistic theory that xenophobia, racism, anti-Semitism, Islamophobia, homophobia, misogyny and so on are slowly decreasing over time (Pinker 2018, ch. 15), people disagree about the role of hate speech and hate speech laws in this. Some people believe that hate speech is both a driver of, and the canary in the coal mine for, prejudice, bigotry and hatred in society and that if hatred is slowly decreasing, then it must be partly thanks to the emergence of hate speech laws (see Delgado 1982; Delgado and Yun 1994b; Delgado and Stefancic 1996). Other people insist, however, that hate speech is itself a kind of necessary safety-valve or pressure-release mechanism, meaning that hate-motivated crime and violence actually decline where people are free to engage in hate speech without threat of legal sanction (Emerson 1963; Heins 1983). Indeed, some

people believe that the very existence of hate speech laws actually makes the problems of intolerance, division and hatred worse (see D'Souza 1991; Harel 1992; Heinze 2006, 2016, ch. 5; Tatchell 2007; Baker 2009; Volokh 2010; Neller 2018) (see Chapter 6 [6.3.3]).

These issues are compounded by the globalisation of hate speech and its effects. When states use legal measures (i.e. domestic hate speech laws) along with various extra-legal measures to combat hate speech, they are reliant on other states to do likewise. If, however, other states fail to do their bit and the problems of intolerance, division and hatred get out of control and escalate into widespread discrimination and violence in those other states, then neighbouring states may end up becoming involved in humanitarian aid, refugee flows, peace keeping responsibilities, and dealing with downturns in economic markets and even conflict itself (Kübler 1998).

Another aspect of the globalisation of hate speech is the rise of Internet-based communication and social networking. Online hate speech in particular is no respecter of state borders. The worry is that if powerful states like America remain implacably opposed to hate speech regulations, or if powerful multinational Internet companies like Twitter, Facebook and YouTube simply have weak enforcement of their own community standards on hate speech, then it no longer matters what the rest of the world does because hate speech will travel between countries through the Internet.

However, this book argues that there are eight reasons to be optimistic that genuine progress can be made, and is being made, on the issue of hate speech and hate speech laws: signs that political communities and the international community as a whole are already moving beyond seemingly entrenched positions and intractable disputes on this issue and are capable of moving even further.

First, governmental authorities, civil society organisations (CSOs), non-governmental organisations (NGOs), intergovernmental organisations, academics and other stakeholders are beginning to commit themselves to more tirelessly and rigorously amassing evidence of the extent of hate speech, its many harmful effects, the different ways of combating hate speech – including through counter-speech, education and different kinds of hate speech laws – and the comparative efficacy of these different sorts of measures. Perhaps only with these efforts will sceptics come to acknowledge that hate speech laws might have a legitimate place within larger bodies of measures designed to combat hate speech. We discuss some of these developments later in this chapter and in the rest of the book.

Second, there is increasing awareness that hate speech does not always or necessarily have to do with feelings, sentiments or attitudes of spite, vindictiveness and intense dislike or hatred toward its targets (Brown 2017a). Building on important insights in the field of hate crime studies (see Jacobs and Potter 1998; Iganski et al. 2011; Citron 2014), there is a growing understanding that hate speech can also come from, and aim at, contempt, fear, anxiety, mistrust, disenfranchisement, alienation, social exclusion or even loneliness (Brown 2017a). Greater understanding of, and even a degree of empathy for, what drives hate speakers opens up new possibilities. Whilst this understanding might not alter some people's view that hate speech is morally problematic and that hate speech laws are necessary, it could help to reduce the tendency or perceived tendency of critics of hate speakers to themselves adopt a posture of spite, vindictiveness and hatred toward hate speakers. Hate propaganda expresses ideas based on the fundamental moral inferiority or lower moral worth or dignity of those being targeted as though it matters less how well or badly their life goes. As such it is unhelpful if critics of people who use hate propaganda express this criticism in a language which suggests or could be read as suggesting that they in turn believe that hate speakers are fundamentally morally inferior or have lower moral worth or dignity. One can believe that

what another person is doing or saying is morally bad and because of that hold them in lesser social esteem or regard measured on ordinary scales of good or bad moral agency, without at the same time thinking that the person is fundamentally morally inferior or has lower moral worth or dignity. Once again, we believe that in order to make progress in these directions more comparative work needs to be done on how the problem of hate speech is viewed in different countries. Increasing this understanding may help raise consciousness about how best to talk about the problem of hate speech without lapsing into the same bad habits. We undertake some of this comparative analysis in Chapter 2.

Third, there is an emerging international norm on hate speech which requires states to use legal prohibition as one measure to combat hate speech alongside a range of other measures including education and counter-speech. The norm has been formalised in a growing body of international hate speech instruments. In Chapter 4 we examine more closely this body of international hard and soft law and explain, using several approaches in international relations theory, why some states do and other states do not agree, sign, ratify and comply with these instruments. Now, it is true that states are sometimes prepared to engage with international treaties on hate speech because of a perception that the problem of hate speech is found in other countries, not their own. But this perception is increasingly being supplanted in many countries with an awareness that hate speech is occurring in their own backyards, offline as well as online. This is due in no small measure to the work of intergovernmental organisations and bodies who monitor compliance and/or adjudicate on individual applications and complaints under international hate speech instruments. Indeed, even states that have historically been resistant to enacting hate speech laws at home have in recent years responded to domestic pressure as well as to external diplomatic criticism and to legal obligations under international hate speech instruments by introducing such laws – Japan being a case we shall discuss later in this chapter and in Chapters 2 [2.8] and 4.

Fourth, it is increasingly recognised that calling on states to make hate speech unlawful does not mean accepting as legitimate every kind of hate speech law and every use of hate speech laws. It is perfectly compatible with affirming the appropriateness of hate speech laws in general to also insist that states should not misuse hate speech laws to suppress political opponents or dissent. Just as international non-governmental organisations (INGOs) have highlighted as a pressing issue of human rights the failure of some states to enact or properly enforce domestic hate speech laws, so INGOs have also identified as another human rights issue the practice of some states misusing or abusing hate speech laws for political ends (see Human Rights Watch 1992a, 1992b, 2005, 2017; ARTICLE 19 2013, 2016; Amnesty International 2018). There is no reason why both sets of concerns cannot be justified. Indeed, this nuanced view on hate speech laws is reflected in the emerging body of international hate speech instruments, especially soft law, to be discussed in Chapter 4 [4.2]. Consider General Recommendation No. 35 on Combating Racist Hate Speech[7] put forward by the UN Committee on the Elimination of Racial Discrimination (CERD) and General Policy Recommendation No. 15 on Combating Hate Speech[8] promulgated by the Council of Europe's human rights monitoring body, the European Commission Against Racism and Intolerance (ECRI). Both recommendations call on states to enact hate speech laws whilst also warning against the abuse of such laws.

Fifth, another important way in which states and the international community as a whole are addressing the globalisation of hate speech is through media and Internet laws. These laws aim to address the problem of weak enforcement by multinational Internet companies of their own community standards and terms of service on hate speech, especially in relation to illegal hate speech. States like Germany are at the vanguard of an international movement

to tackle the globalisation of hate speech by adopting media and Internet laws that hold multinational Internet companies accountable for failing to remove illegal hate speech content from their platforms, services and websites. We shall look at these efforts in Chapter 2 [2.11].

Sixth, more attention is being devoted to critically evaluating and empirically challenging the credibility of some of the standard political objections to domestic hate speech laws. An important part of this project involves detailed investigation into why given states have ended up with the particular hate speech laws they have and what functions these laws continue to serve. This is a vital first step to evaluating these laws as being either fit or unfit for purpose. We make a small contribution to this effort in Chapter 3 where we investigate the reasons behind the introduction of stirring up religious hatred offences in England and Wales.

Another important part of this rehabilitation project is to debunk some of the 'bad' political arguments used against hate speech laws – arguments which in one way or another misconstrue the nature, purpose or grounds of such laws or else simply get the facts wrong. One common argument is that hate speech laws constitute a slippery slope to massive censorship. Another is that despite being well-intentioned, hate speech laws are actually responsible for exacerbating intolerance, division and hatred between sections of society. We point out the significant flaws in these arguments, and other 'bad' arguments, in Chapters 5 and 6 respectively.

Seventh, parliamentary committees, courts, intergovernmental organisations, INGOs and other stakeholders are paying increasing attention to political figures not merely as decision-makers concerning the enactment or repeal of hate speech laws but also as hate speakers. Nevertheless, there is an urgent need for better understanding of how the power, authority and influence of political figures can amplify the meaning and effect of their hate speaking and why this can be morally problematic. We aim to make a significant contribution to this understanding in Chapter 7.

Finally, it is clear that increasing pressure is being placed on political figures on both sides of the right-left access, both from the aforementioned stakeholders and from the media, ordinary citizens and other politicians, to recognise the special responsibilities they have, as holders of hard and soft power, to set an example by refraining from the use of hate speech. Moreover, legislators in many countries are taking more seriously the need for legal or quasi-legal measures to combat the use of hate speech by political figures, including both parliamentarians and candidates. In Chapter 8 we attempt to move this debate forward by surveying and assessing the pros and cons of a range of possible measures. In the end we argue that both legislators and the judiciary should no longer be ambivalent about applying existing hate speech laws to political figures when they use illegal hate speech, and should limit the scope of parliamentary privilege to make this possible. We also argue that the content and enforcement of party, parliamentary and election codes of conduct should be strengthened to enable sanctions against political figures who use soft hate speech.

The remainder of this chapter has the following structure. In the next section we start the process of fleshing out the connections between hate speech and politics. We outline how some, but not all, uses of hate speech are politically motivated. We also examine more closely some of the many ways in which the terms 'hate speech' and 'so-called hate speech' have themselves become highly political and politicised. And we look at some political disputes concerning what to do about hate speech.

Following on from this, we address various forms of cynicism about the problem of hate speech. One idea is when both sides in a debate about hate speech accuse the other of using language as an extension of power, then there is nothing that can, or should, be done to tackle hate speech. However, we argue that recognising the relationship between hate speech and power is perfectly compatible with making rational determinations of where the

most extreme and troubling exercises of power are located, including the ways in which hate speakers use language to discriminate, oppress and subordinate marginalised and vulnerable groups in society. In addition, we tackle Stanley Fish's suggestion that the real problem of hate speech is in calling it 'a problem', which we reject on several fronts. Next we challenge head-on the idea that because people disagree about what counts as hate speech it should never be prohibited. We argue that disagreement is not the same as mutual incomprehension. We also address attempts in the literature to define the legal concept hate speech and argue that it is important not to forget context in the process. And we dispute the cynical position that the problem of hate speech is simply in the eye of beholder and that there is no genuine harm in hate speech. We counter the myth of harmlessness with an overview of the many kinds of harms that can be done by hate speech.

After that we try to map out some of the main sites in which problems of hate speech present or locate themselves. The sites of the problem include: individuals, groups, society, technology, law, politics and international relations.

Finally, we address some remaining preliminaries. We explain the different methodologies employed in the book. We briefly survey the canon of literature on hate speech to which we seek to make a contribution (to challenge, develop or simply move beyond), as well as the wider topics of academic study that our investigation will overlap with in places. And we provide short summaries of each of the chapters.

1.2 Hate speech and politics

At the start of this chapter we stated rather enigmatically that hate speech is a political and politicised issue. But what does this mean exactly?

1.2.1 Politically motivated hate speech

Some people might be tempted to think that all hate speech is itself political in the sense that it emanates from political motives or is always oriented toward political goals and agendas. Now it is undeniably true that for some, perhaps many, hate speakers the drive is political in one way or another. They seek: to articulate ideological beliefs about differences between human beings, to affirm their own identity which they believe to be under threat under the current political settlement as a rallying cry to like-minded members of the political community, to maintain social hierarchies, to threaten or terrorise an outgroup so as to gain control over them or else cause them to exit society, to call attention to the perceived injustice of their reduced circumstances, to justify a policy agenda, to dehumanise others so as to make harmful policies seem more acceptable, and so on.

Much the same variety in motives for using hate speech can be found among political figures, those seeking political power as much as those holding it. Because politicians are 'of the people' it is inevitable that the use of hate speech, especially in its implicit or coded forms, can be sometimes akin to unconscious bias. Then again, for many others the use is more deliberate. It may be a function of political figures: wanting to seem 'normal', looking tough, responding to political rivalry, talking to the base, playing to people's fears, a race to the bottom, diversionary tactics, shoring up political power, grandstanding, cutting through the white noise, being or playing the faithful delegate, dirty hands, political conviction, making voters feel better about their own prejudices, dissent, becoming a martyr to a cause (see Brown 2019b).

Furthermore, a great deal of hate speech expresses or connotes ideas that have political meaning in the context in which they are expressed. Hate propaganda, for example, expresses ideas based on the fundamental moral inferiority of certain races, of people with

a particular gender, of people with certain sexual orientations, of people with particular mental or physical disabilities and so on. Such speech is at its core about the nature of what it means to be human or a good human, but invariably this also connects with broader ideas about the public good and about good societies. More often than not the function of using hate propaganda to express these sorts of ideas is to say something about a given political context: about how a particular society should be organised (e.g. racial segregation), about how a government should treat a certain group of people (e.g. census taking, deportation, eradication) and about a utopian vision of what the future direction of a certain country should be (e.g. a white civilization). Similarly, the use of slurs, negative stereotypes, group libels and false rumours about vulnerable groups or minorities often has both a political significance and a political purpose in societies where particular minorities have historically faced structural injustice, oppression and discrimination.

However, it would be wrong to assume that all hate speech is politically motivated in the aforementioned senses. Whilst it is true that to use a racial slur, for example, is rarely *just* to use a racial slur, some motivations seem more human than distinctively political. For some people using a racial slur is not a political statement but rather: an instinctive outpouring or expression of fear or loss, a displacement of aggression, getting a kick out of humiliating others or feeling powerful, an attempt to prove oneself, a desire to fit in within a social group, a cry for help of the lonely or alienated, self-promotion, even just a way of increasing clickbait revenue.

1.2.2 The use of the term 'hate speech' as a political or politicised act

Another way of coming at the relationship between hate speech and politics is to point out that the use of the term 'hate speech' has become highly political or politicised. On the one hand, critics of anti-hate activists sometimes accuse them of grossly misrepresenting, taking out of context, exaggerating or even inventing claims made by those they (the anti-hate activists) condemn as 'hate speakers' for political reasons, without bothering to actually engage with the substance of the opinions being expressed. The politics of the term 'hate speech', according to these critics, is simply the politics of a label used to silence speech that certain sections of society do not like or do not want to hear, despite never bothering to truly listen. From this political worldview, the terms 'hate speech' and 'hate propaganda' are themselves propaganda words.

On the other hand, anti-hate activists claim that it is their critics who grossly misrepresent their views and intentions, and who turn them into straw men or else demonise them, not least by labelling them as 'political correctness zealots' or 'thought-crime police'. They argue that phrases like 'so-called hate speech' and ' "hate speech" is a propaganda word' are themselves political. These phrases are born of ideology, propaganda and the ambitions of conservative political movements.

In fact, taking a view on what hate speech is or the problem of hate speech, one way or the other, has become an important and prominent way of identifying or affiliating with political positions, parties or ideologies. This is true in many of the country contexts we will discuss in Chapter 2. The idea that hate speech and hate crime are serious social problems that need to be addressed is a typical belief among liberal progressives and those on the political left. In Western popular culture, speaking out against hate speech, especially the hate speech of Donald Trump, has become a cause celebre for liberal politics. Consider what the Harry Potter author, J. K. Rowling, tweeted after Trump won the 2016 presidential election.

> We stand together. We stick up for the vulnerable. We challenge bigots. We don't let hate speech become normalized. We hold the line.[9]

The phrase 'We hold the line' has a militaristic connotation, as though liberals are facing off against an onslaught or offensive push from the enemy.

Conversely, the idea that so-called hate speech is just a trapping of the ideology of political correctness is a typical belief among civil libertarians, conservatives and those on the political right. During the 2016 election campaign part of Trump's rhetorical strategy was to construct an image of himself as a popularist maverick and anti-establishment candidate by suggesting that 'the big problem this country has is being politically correct'[10] and by declaring with pride, 'I refuse to be politically correct'.[11] Take the idea that political correctness, including restrictions on hate speech, has been weaponised by legislators, the media, campaign groups and public intellectuals against ordinary Americans, as a way of silencing their dissent against the liberal establishment. This idea has become for some people on the right not merely a peripheral idea but rather a core belief or tenet. It has become a defining idea among people wishing to self-identity as conservatives or of the political right and something they publicly espouse at times to flag up or affirm their credentials as such. It has become a matter of self-expression or self-realisation in that sense. During the 2016 campaign some Trump supporters, for example, attended his campaign rallies wearing T-shirts that read, 'If you're afraid to offend, you can't be honest, #TRUMP2016'.

1.2.3 Political disputes concerning what to do about hate speech

There is another, related sense in which hate speech is or has become highly politicised. Protagonists on all sides of the mainstream debate about what to do about the problem of hate speech, including whether or not to ban it, 'often take vibrant political stands' (Heinze 2016, 8). Trump, for example, has suggested that Internet companies' community standards on hate speech are biased against social conservatives and people on the political right, albeit ironically he has freely utilised the medium of Twitter to voice that accusation.[12] However, two pressing questions are, first, whether the difference between hate speakers and opponents of hate speech always maps squarely onto the right-left axis or even the nationalist-cosmopolitan axis, and, second, whether, even if it turns out that a given hate speech regulation has discriminatory effects, this really does constitute unfair discrimination. We shall tackle these questions in Chapter 6 [6.3.1].

Campus speech codes, which typically ban discriminatory harassment (a form of hate speech), have been another flashpoint in American politics. Campus speech codes have not merely been subject to strict scrutiny by the courts – we shall discuss in Chapter 2 [2.7] the many instances in which US courts have struck down campus speech codes as unconstitutional – but they have also come under fire for allegedly unfairly targeting and disproportionately chilling the speech of students who hold socially conservative or right-wing views (see Jesse 2018).

Trump has also weighed into the debate about anti-conservative bias on university and college campuses. Speaking to the Conservative Political Action Conference in March 2019, for example, Trump invited onto the stage Hayden Williams, a conservative activist who, whilst trying to recruit students for a conservative group, had been assaulted by a student at the University of California, Berkeley.

> There's a young gentleman – I turned on my television the other day, and I saw somebody that was violently punched in the face, violently punched, and I said, 'That's disgusting', by a bully, I'd like to do a lot of things, but of course we would never do that, because if I ever said violence they would say 'Donald Trump attacked', no no, just for

> the media, I'm sure he's a lovely young man, just had a little temper tantrum. [. . .] The man's name is Hayden Williams. [. . .] I know your lawyer, she is a great lawyer. Just tell her to do me a favour, sue him, but he's probably got nothing, but sue him forever, but sue the college, the university. And maybe sue the state. Ladies and gentlemen, he took a hard punch in the face, for all of us. Remember that. [. . .] Today, I'm proud to announce that I will be very soon signing an executive order requiring colleges and universities to support free speech if they want federal research dollars. If they want our dollars, and we give it to them by the billions, they've got to allow people like Hayden and many great young people, and old people, to speak. Free speech.[13]

Rather than recommending that conservative or right-wing students themselves make use of campus speech codes insofar as they believe they have been subject to discriminatory harassment or violence based on their political views – and we shall consider whether political views should be a protected characteristic in Chapter 5 [5.5.3] – Trump's advice to students is to sue universities for failing to support free speech.

However, appealing to free speech is rarely an ideologically neutral move. Both people who fear being prosecuted under hate speech laws and academic scholars who do not engage in hate speech but who defend the rights of people who do engage in hate speech invariably draw on their own ideological commitments when interpreting what free speech is and explaining why it is valuable. For example, some scholars who vehemently oppose hate speech bans, especially campus speech codes, appeal to distinctively liberal values such as ethical individualism (see Dworkin 1996, 200–1, 2011, 372–3, 2012, 342–3; Shibley 2015). Others invoke conservative ideas of rugged individualism or self-reliance (see D'Souza 1991, 242, 2018) or having a healthy suspicion of government (see Gellman 1991, 391–2). Meanwhile, people who deny that individuals have an absolute right to engage in hate speech also draw on their own ideological commitments to justify hate speech bans. Sometimes the arguments invoke ideals of equal protection and the rights of people not to be targets of discriminatory harassment especially in contexts like the workplace and university campuses (see Lawrence 1990, 1992; Matsuda 1989b; Matsuda et al. 1993; Delgado and Stefancic 2004). These arguments (on both sides) are ultimately rooted in different political ideologies, each presenting interdependent clusters of interpretations of key interests, rights, principles and values, such as to do with freedom, dignity, autonomy, security, equality, mutual respect, human rights, democracy and legitimacy. We shall seek to trace out these different ideological commitments in Chapter 6 [6.5.2]. We shall argue that the problem lies not in the fact of making arguments based on deeper ideological commitments but in the act of accusing only the other side of having such commitments.

However, it is also important to recognise that disputes about hate speech laws are not merely dependent on political ideologies but also sometimes speak directly to distinctively political or democratic values. For example, some scholars argue that hate speech laws pose a substantial threat to democracy and democratic citizenship (Heinze 2016) or to democratic legitimacy itself (see Weinstein 1999, 2001, 2009, 2011, 2017a, 2017b). But others defend hate speech laws on the grounds that they protect public assurance of civic dignity (see Waldron 2010, 2012a, 2012b) or that they safeguard real access to public discourse and democratic decisionmaking and even normative political legitimacy (see Fiss 1996; Parekh 2005–6; Brown 2015, ch. 7, 2017d; Waldron 2017; Reid 2019). We shall revisit some of these arguments in Chapters 7 and 8, albeit we shall focus on the particular issue of laws and rules restricting political figures' use of hate speech.

1.3 So what is the real problem of hate speech?

Reflecting on the political and politicised nature of debates about hate speech sometimes inspires a world-weary view that the problem of hate speech is *merely* a political problem. On this view, there is no actual problem of hate speech, there is only a political problem involving a dispute between people with competing interests and ideological agendas. In this section we challenge such cynicism directly.

1.3.1 Understanding what the concept hate speech reveals about power

The American sociologist Charles Wright Mills once said of 'intellectual work' that '[w]ith ideas one can conceal or expose the holders of power' (Wright Mills 2008, 135). There is no doubt that this is how critical race theorists like Mari Matsudi, Charles Lawrence and Richard Delgado first intended their use of the term 'hate speech', namely, to expose the holders of power (see Matsuda 1989b; Lawrence 1990; Delgado 1991, 2000). In other words, they wanted a term that could be used to characterise, or simply as shorthand for, certain modes of speech typically used by the powerful to oppress, subordinate and dominate the less powerful and the vulnerable. From this perspective, hate speech is not merely an expression of power but is a tool or extension of that power and also a means of its reproduction. 'Hate speech today is a central weapon in the struggle by the empowered to maintain their position in the face of formerly subjugated groups clamoring for change' (Delgado and Yun 1995, 1298).

At the same time, however, Wright Mills also continues more darkly: 'And with ideas of more hypnotic though frivolous shape, one can divert attention from problems of power and authority and social reality in general' (Wright Mills 2008, 135). There is no doubting the fact that this is how some conservatives and people on the political right view contemporary uses of the term 'hate speech', that is, a way of some people seeking to divert attention from social reality. In their minds, the term 'hate speech' fits into a broader category of accusatory and pejorative terminology, such as 'unconscious bias', 'white privilege', 'bigoted' and 'racist'. Intellectual elites use such terminology simply as an excuse to summarily dismiss or sweep under the carpet the legitimate concerns and fears of certain sections of society about the effects of globalisation and large-scale immigration on their economic position, and threats to their culture and traditional family values from progressivism. From this perspective, the real problem of hate speech is the utilisation of this idea by some people to close down debate and chill the speech of socially conservative and right-wing politicians and their supporters. The idea of hate speech, in other words, is not just an affront to the very universal values like free speech that liberals purport to endorse; it is also one of the newest weapons in the arsenal of anti-conservativism and simply an extension of the dictatorship of a perverted version of liberalism. To give a concrete illustration, political activists who have been publicly accused of engaging in Holocaust denial and anti-Semitism have attacked the very use of the idea of hate speech by liberals. 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' (cited in Brown 2017b, 572).

Reflecting on this power struggle over the term 'hate speech', we believe there are four significant dangers to look out for. The first is how quickly the struggle leads to a dead end. After all, Rivero's attack on his critics looks very close to the claim, 'I protest at your use of the term "hate speech" to shut me up, and in return I use the term "so-called hate speech" to shut you up'.

Second, whilst liberal progressives should acknowledge that social conservatives and people on the right of politics may have legitimate grievances about the use of the label 'hate speech' in some cases, there can be a danger in over-accommodation. People who are worried about hate speech should not be afraid to be forceful and determined but also consistent and fair in explaining why they have grounds to make criticisms. Just because not all discourse about immigration, for example, constitutes xenophobic, racist or Islamophobic hate speech, it does not mean that no such discourse is hate speech.

Third, clearly there are people who say, 'The less we fuss and legislate about hate speech, the better things might get'. But this raises some important questions that should not be ignored. What reason do they have to think change will happen simply of its own accord? Why are they advocating not making a fuss about hate speech? What is it about the status quo they favour? Do they have positions of power and privilege in the status quo they are keen to protect? In other words, there is simply no getting away from the fact that sometimes people, including political figures, use hate speech for certain political purposes and have an interest in not being challenged or prevented from doing so. Thus, it is no coincidence that sometimes people use hate speech against groups whilst at the same time seeking to obstruct the granting of basic civil or human rights to those very groups. Consider the case of the Canadian member of parliament (MP) who claimed that extending certain human rights to members of the LGBT community would be akin to granting protection to 'bestialists, paedophiles and necrophiles' (Faulkner 2006–7, 73). Surely to deny the existence of these sorts of uses of language is itself to attempt to divert attention from or mask the social realities of politics itself. Arguably we need concepts like hate speech to express these realities. In other words, '[a]mong the many jobs that are, or might be, performed by the concept hate speech [. . .] [is] highlighting forms of speech that it is believed disproportionately harm already disadvantaged or victimised members of society' (Brown 2017b, 569).

Fourth, many people believe that it would be problematic – perhaps the greatest problem of all – to live in a world in which forms of discrimination, oppression and subordination were present but concepts like hate speech which challenge the status quo did not exist, could not exist or were simply not permitted to exist. This would be a world in which people in power would have a complete stranglehold, such that they could suppress the very concepts needed to unmask their power and the ways they exercise it. Yet there is also a danger that people with these beliefs at the same time underestimate or fail to fully acknowledge that these same concepts can also be abused or misused by authoritarian governments to suppress political dissent. By highlighting one problem it is important not to ignore other, related problems. Otherwise people who disagree may suspect that the first problem is illusory. We look at some examples of oppressive or abusive uses of hate speech laws by authoritarian governments in Chapter 2.

1.3.2 The real problem of hate speech is not simply in calling it 'a problem'

Some scholars object not so much to the concept hate speech but to calling it 'a problem'. On this view, the real problem of hate speech lies in the tendency to classify it as a problem. This way of thinking is perhaps best exemplified in the following passage taken from the work of Stanley Fish.

> Consider, for example, what has often been called the "problem" of hate speech [. . .]. I put "problem" in quotes so as to flag it as a noninnocent usage. The problem with "problem" is that, unqualified, it means "problem for everyone," or a problem universally; as in the problems [. . .] of hunger and disease. Hate speech, however, is not such

> a problem; rather, it is the pejorative designation by one party of the way of thinking and talking central to the beliefs and agenda of another party. Moreover, not only is it a mistake to term hate speech a problem, and thus to imply that a cure for it may someday be found in a pill or a book whose consumption will render any reader morally healthy, it is a mistake with consequences. It gets in the way of thinking about strategies for dealing with that which you regard as dangerous and a source evil. If the evil is given no particular location but is regarded more or less as a virus that mysteriously affects some people and leaves others uninfected, you will think in terms of remedies or, in cases where the disease is too far gone, of quarantine. But if the evil is given a location in a worldview you despise and fear – not an irrational worldview (calling speech you loathe "irrational" is another form of universalizing and trivializing) but a view equipped with reasons, evidence, and authorities you reject and find truly harmful – you will think in agonistic and political terms and begin to figure out how you can stigmatize, oppress, and in general get the better of an enemy.
>
> (Fish 2001, 148–9)

However, we take issue with Fish on four points. First, arguably hate speech is a universal problem, and in at least two different senses. For one thing, if the problem does really exist, then there are reasons to believe that it exists in virtually every country around the world and has existed perhaps for as long as human beings have had language to describe certain people (tribes) as 'other'. The mere fact that some sections of society more often find themselves falling foul of a problem – such as destitute people and the problem of hunger – does not mean that the problem is not universal. For another thing, hate speech is a problem both for people who think it really exists and for people who think it is simply a liberal construct. After all, so long as there are people who denounce the use of hate speech, who refuse to be addressed with hate speech, who will not stand by and let others be victimised by hate speech and who call for hate speech laws, then that very fact alone means that people who want to use the speech in question are going to have a problem. Nevertheless, if what Fish is actually saying here is that there is a problem of hate speech and a problem of anti-hate speech, then he is surely right. And we can use the term 'the problem of hate speech' to mean both. This reflects the fact that both hate speech and responses to hate speech can be a problem from whichever perspective you are looking at it from.

Second, to regard something as a 'problem' does not strictly entail the assumption that it has a solution or is on par with a disease with a potential cure. This speaks to a wider point that some problems are tragic in the sense that no response is without at least some cost to at least some of the people concerned. When it comes to hate speech, we are damned if we ban it and perhaps damned if we do not (see also Brown 2015, ch. 9). The fact that hate speech is a problem to which no response (banning or not banning) leaves everyone better off and nobody worse off means only that it is a problem we would all prefer not to have, not that it is an imaginary problem. Indeed, the idea that hate speech is a problem we would rather not have is probably the one thing that virtually everyone can agree on, even those who think it is a myth.

Third, simply using the phrase 'the problem of hate speech' is not the same as seeing this as an abstract, transcendental or unencumbered problem. On the contrary, most social problems are disproportionately problems for certain sections of society. But, even more importantly, it cannot be simply assumed that the problem of hate speech has always been and will forever be a problem for some fixed set of groups in society, such as people with dark skin, Jews or homosexuals, for instance. And just as we cannot always know in advance who the 'victims' of hate speech will be, so we cannot presume to know who the hate speakers will be.

We cannot take for granted that hate speakers are always white, male, socially conservative right-wingers. In the UK, for example, the EDL has sought to unite a broad coalition of people with Islamophobic views, welcoming people who are themselves members of racial, ethnic or religious minority groups. Indeed, the EDL has prided itself on being not merely a multi-ethnic movement but also vehemently anti-Nazi (see Allen 2011). By the same token, the leader of the Labour Party, Jeremy Corbyn, has been dogged by accusations of his failure to deal adequately with a problem of anti-Semitism in the party, even though it has a general and longstanding tradition of anti-racism (Sabbagh 2018a). Towards the end of 2018 Corbyn was himself accused of making anti-Semitic comments when a 2013 video clip emerged of him accusing a group of British Zionists of having 'no sense of English irony', the implication of his words being (according to some) that because these people were Jews they were somehow not really English (BBC 2018c). Therefore, it may be correct to talk in general terms of 'the problem of hate speech' because doing so shows an apposite awareness that hate speakers can have almost any ethnicity or political creed.

Finally, Fish seems to think that people who are opposed to hate speech should embrace the idea that hate speech emanates from 'a worldview [they] despise and fear' because this will encourage them to 'think in agonistic and political terms and begin to figure out how [they] can stigmatize, oppress, and in general get the better of an enemy'. However, we believe that part of the problem of hate speech is precisely the tendency or perceived tendency of critics of hate speakers to themselves adopt a posture of spite, vindictiveness and hatred toward hate speakers. Arguably this posture carries the risk of precipitating a defensive reaction from hate speakers. In some cases it simply prompts the retort, 'No, you're a hate speaker!' Not all political problems can be solved simply by turning up the dial on agonism. Maybe the problem of hate speech is among those problems best addressed in a spirit of collaboration, consensus and compromise. In practical terms this may require people who are worried about hate speech to think about how they can work with hate speakers to figure out what it would take for all sides to recognise it as genuinely harmful, and for hate speakers and people who defend the free speech rights of hate speakers to think about how they can work with people worried about hate speech to figure out what it would take for all sides to recognise the costs of badly drafted hate speech laws.

This does not mean that people who are worried about hate speech should no longer call for it to be banned, however. On the contrary, they should. Rather it means they should articulate their grounds without resorting to name calling or making crude assumptions that all hate speakers are socially conservative or right-wingers or that all hate speakers are driven by the most intense hatred, bigotry and contempt for the targets of hate speech. Part of the power of hate speech lies in the fact that it is used so commonly and by all sorts of different and otherwise law-abiding people. Sometimes it is correct to despise the hate speech, not the hate speaker. When it comes to the stronger forms of moral reproof and denunciation, perhaps these should be reserved for hate speakers who are responsible for promoting, legitimising or normalising hate speech and for influencing others to join in the practice of hate speaking. In Chapters 7 and 8 we shall focus on one such constituency of hate speakers: political figures.

1.3.3 So what that people disagree about what to count as hate speech?

The idea that authorities should not attempt to ban hate speech because (allegedly) people do not agree about what counts as hate speech has at times been mooted not only by American free speech scholars (see Strossen 2018) but also by the US government. Speaking at an international conference on hate speech in Budapest in April 2006, for example, the then

US ambassador to the Organization for Security and Cooperation in Europe (OSCE), Julie Finley, put the idea in the following terms.

> The title of this conference seems to imply that there is a well-defined category of "hate speech" and that we all agree on what it includes. Yet the reality is that different people with different perspectives and life experiences will inevitably come to different conclusions. Who will define it? How can we ensure that efforts to restrict "hate speech" don't in fact turn into a tyranny of the majority over minority voices? [. . .] Americans throughout history have chosen not to give up our freedom of expression. We fear censorship much more than we fear offensive speech. [. . .] Efforts to restrict hate speech represent a clear and present danger to robust political debate. Once we start down the slippery slope, trying to define a nebulous term like "hate speech," we are heading for the potential for abuse.[14]

We reject this reasoning. It is one thing to say that drafting hate speech laws is not an easy task because people disagree about what to count as hate speech. Indeed, results from the 2017 Cato Institute/YouGov public opinion survey suggest that Americans overwhelmingly agree (82 per cent) that 'it would be hard to ban hate speech because people can't agree what speech is hateful', albeit agreement is 87 per cent among white Americans but only 59 per cent among African Americans (Cato Institute/YouGov 2017, 10). But it is quite another thing to suggest that just because people disagree about what to count as hate speech no attempt whatsoever should be made to ban hate speech. It being hard or difficult to do something should not preclude the effort to try, if it is the right thing to do. Indeed, simply recognising the truth that 'different people with different perspectives and life experiences will inevitably come to different conclusions' need not, and should not, be used to justify the view that hate speech is merely a subcategory of offensive speech or even a crude shorthand for 'speech that we hate' (see Brown 2017a, 447–9).

Indeed, Ambassador Finley cannot have it both ways. She cannot in one breath say that because people disagree about what to count as hate speech it should not be banned and then in the next breath say the following.

> I am not saying that the U.S. condones hate speech. Quite the opposite. We all need to speak out against expressions of intolerance and hatred whenever they occur. We need to educate people – especially young people – so that they do not develop intolerance or hatred toward any group.

If hate speech should not be banned because its meaning is contested, then on what basis can the US state claim to not condone it or to support extra-judicial measures against it like education or counter-speech? What is the 'it' being referred to?

Now we accept as a technical point that at least when it comes to the common or general concept hate speech it is by no means obvious that this is the sort of concept that can be defined in terms of a set of necessary and sufficient conditions (see Brown 2017a). But that is true of many other family resemblances concepts (Brown 2017b), and it is not a reason in itself not to prohibit some things that are part of the family. No doubt there is much political disagreement about what concepts like corruption and collusion mean, but that does not mean they should be legalised.

Importantly, it would seem that the term 'hate speech' does have a broad set of connotations that people roughly understand, whether they defend or oppose hate speech bans. Despairing counsel that declares the term 'hate speech' either vacuous or hopelessly contested is (wilfully?) ignorant of the fact that people around the world use the term all the

time – or equivalent terms if not direct translations – and have at least some general impression of what everyone is talking about. There is not mass confusion or mutual incomprehension when the term 'hate speech' is used by people with different racial, religious, class, cultural, political, linguistic or national backgrounds, for example. The same goes for the term 'hate crime'. People have a handle on what is at stake or a ballpark impression of the sorts of things the term might refer to. They have this minimal understanding even if they cannot put their finger on the essence of hate speech, that is, a set of necessary and sufficient conditions that would enable everyone to write more or less identical lists of every example.

In ordinary discourse the term 'hate speech' has become an umbrella term, as well as an opaque idiom, with multiple meanings covering a heterogeneous collection of expressive phenomena (see Delgado and Stefancic 2004, ch. 1; Brown 2017a). Yet even though the things to which it applies do not share one thing in common, they have family resemblances or a patchwork of likenesses (Brown 2017b). When speakers use abusive slurs, negative stereotypes, derogatory remarks, dehumanising images, mocking imitations, claims about lower fundamental moral worth or dignity, false rumours, group libels, threatening words or behaviour, denials or glorifications of atrocities, all based on the race, ethnicity, religion, citizenship status, sexual orientation, gender identity, disability or some other protected characteristic of the targeted person or group of persons, especially in circumstances where this speech is intended or likely to rank as inferior, humiliate, intimidate, stir up hatred towards or promote or incite discrimination or violence against these persons or groups of persons, and in turn could lead to psychological or dignitary harms to the individual, to public disturbances or breaches of the peace, to a worsening of social cohesion or community relations, to a breakdown in intercultural communication, to undermining public assurances of equal standing in society, or to a reduction in real access to public discourse and democratic decisionmaking, then most people would recognise this as being in the terrain of the general concept hate speech (see Brown 2017b).

No doubt there are many subtle and important distinctions to be drawn between the members of this heterogeneous collection of expressive phenomena. For example, whereas abusive slurs, negative stereotypes and derogatory remarks typically involve judging members of a group to be criminal, immoral, despicable, animalistic or in some other way less estimable on the everyday scale of moral praise and blame, arguably hate propaganda is about ranking other people or groups of people as in some sense lacking fundamental moral worth or dignity. In the case of hate propaganda the suggestion is not merely that certain people are immoral or despicable, it is that the success and failure of their lives is less important or does not matter, morally speaking. The implication is that certain people lack fundamental moral worth and have mere value in the way that non-human animals or inanimate objects might have more or less value. Hate propaganda says that if certain people were exterminated, there would be no loss of net moral worth in the universe, akin to the moral insignificance of exterminating cockroaches (in their eyes). Such hate speech is doubly problematic in terms of the assumptions it makes about the lack of fundamental moral worth of both certain kinds of human beings and non-human animals. Understandably, hate propaganda, along with Holocaust denial, has been of particular concern in countries home to people impacted by the Holocaust. It has also been highlighted in international hate speech instruments, most notably in Art. 4 of the International Convention on the Elimination of Racial Discrimination (ICERD),[15] which calls on all states to 'declare an offence punishable by law all dissemination of ideas based on racial superiority'. Using hate propaganda is one of the ways in which people incite hatred or bring people into contempt. But it is not the only way. Stirring up hatred can also be done through the use of abusive slurs, negative

stereotypes and derogatory remarks which judge other people as lesser but may fall short of ranking them as lacking fundamental moral worth or dignity.

Nevertheless, from the perspective of this book, the pressing question is less whether the term 'hate speech' can be defined – and there are many competing definitions offered up in the literature (see Brown 2017a) – but whether a given definition can fulfil certain justifiable purposes (see also Post 2012, 31). And it does not take long to realise that there is as much reasonable disagreement about what purposes the concept hate speech should serve as there is about what it means (see Brown 2017b, 567–73, esp. 569). In other words, because people disagree about what the concept of hate speech is supposed to do, that is, its core function or point and purpose, they disagree about what its best conception is.[16]

But clearly one function the concept of hate speech can and does serve is legal: '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 publicly acceptable speech' (Brown 2017b, 569). And, when it comes to the legal concept hate speech in particular (as opposed to the general concept[17]), it is plain wrong to suggest that there is nobody to define hate speech. When Ambassador Finley asks, 'Who will define it?' the answer is that, at least when it comes to the legal concept hate speech, it will be defined in law by legislatures, courts and intergovernmental organisations of various kinds. They will no doubt take into account aspects of the ordinary concept hate speech, but they will also trim, polish and refine the definition to suit the legal purposes sought. More importantly, legislatures, courts and organisations will be judged as tyrannical or not depending on what fist they make of defining hate speech, as they are judged in how they define all legal concepts. If they are tyrannical, then they, not the concept, are to blame.

1.3.4 Defining the legal concept hate speech, whilst not forgetting context

The legal concept hate speech is partly defined, whether explicitly or implicitly, in domestic law and in international hate speech instruments, just like legions of other human rights concepts. Moreover, just as the international community comes together to agree hard law (treaties) relating to hate speech, so intergovernmental organisations of various kinds make soft law (decisions, individual communications and recommendations) relating to hate speech (See Chapter [4.2]). Thus, Ambassador Finley's comments speak perhaps more to the US view on the importance of internationalism than they do to the technical issue of defining the legal concept hate speech.

There is also potentially tremendous value in philosophers, legal scholars and political scientists seeking to flesh out the contours of something like an ideal legal concept hate speech. This involves going beyond simply how the term 'hate speech' is defined in particular legal documents to how it should be defined or what the best version of the concept could be (see Brown 2017a, 422–3). Some of this work sets forth core elements or features of what an ideal legal concept hate speech would look like, such as in terms of the sorts of harmful or discriminatory speech acts it would make unlawful (see Matsuda 1989b, 1998, 2013; Tsesis 2002; Delgado and Stefancic 2004; Cohen-Almagor 2009; Langton 2012; Langton et al. 2012; Gelber 2012, 2017, 2019). Much of this scholarship owes an intellectual debt to critical theory in general and to feminism and subordination theory in particular.

That being said, scholarly attempts at constructing an ideal legal concept hate speech have tended to err on the side of claims to universality. Typically the aim, consciously or unconsciously, has been to define the legal concept in general enough terms to be applicable throughout the world. However, we believe that any plausible or minimally adequate

account of the legal concept hate speech should give space for how this concept is shaped by different country contexts. This is important for many reasons, not least of which being dewesternising and decolonising hate speech law.

Thus, in Chapter 2 we shall argue that contextualising hate speech is not simply a matter of recognising that different countries understand the problem of hate speech differently, although this is important in itself. It is also about acknowledging the fact that different countries may seek to define the legal concept hate speech differently and that some degree of contextual differences can be perfectly legitimate. In other words, the project of defining the legal concept hate speech should be sufficiently sensitive to the particular historical, political, legal, economic, social and cultural circumstances of given countries. This is partly why international human rights courts, such as the European Court of Human Rights (ECtHR), provide a 'margin of appreciation' for countries to read down their obligations under international law in ways that are compatible with their own domestic hate speech laws, for example. It is also why intergovernmental organisations, such as the CERD, make general recommendations that leave scope for interpretation at the country level.

1.3.5 Acknowledging the real harms of hate speech, in spite of the politics

Earlier we cited comments made by US Ambassador Julie Finley from 2006 in which she lumps hate speech into the broader category of offensive speech (thus herself defining its nature). This is a common rhetorical strategy for those opposed to banning hate speech. Portraying hate speech as merely offensive speech can make it easier to justify not using legal measures, among other sorts of measures, to combat it.[18] But this strategy is mistaken because hate speech is not the same as merely offensive speech. Moreover, the strategy is founded on a dangerous myth: the myth of harmlessness.

Hate speech is, whatever else it happens to be (dog whistle, political football, alarm bell, etc.), a family of expressive phenomena that can be, and often is, genuinely harmful. Hate speech can pose a real danger to public order and security, it can reduce social harmony and community cohesion, it can damage the psychological well-being of victims, it can oppress or subordinate its targets, it can chip away at the autonomy of the audience, it can diminish access to truth, knowledge and self-realization, it can violate human dignity, it can misrecognise cultural identity, it can transgress norms of intercultural dialogue, it can undermine public assurances that persons are members of society in good standing, it can impede real access to participation in public discourse and democratic decisionmaking and it can call into question the political legitimacy of regimes of rules and laws (see also Matsuda 1989b; Lawrence 1990, 1992; Matsuda et al. 1993; Altman 1993; Nielsen 2002; Tsesis 2002, 2017; Delgado and Stefancic 2004; Parekh 2005–6, 2006, 2012; Heyman 2008; Maitra 2012; McGowan 2012; Langton et al. 2012; Waldron 2012a, 2012b, 2017; Brown 2015, 2017d, 2018b; Gelber and McNamara 2016; Gelber 2017, 2019).

Of course, there is still much more evidence that needs to be collected, and more debate to be had about the evidence that has already been collected. This goes for evidence of the putative damage to people's psychological well-being (see Brown 2015, 49–58), evidence for the putative connection between hate speech and acts of violence (see Hinkle 2015; Brown 2015, 66–75) and evidence for the putative silencing effects of hate speech, specifically the claim that hate speech can undermine equal access to participation in public discourse and the formation of democratic public opinion (see Weinstein 2001, 2017a, 2017b; Brown 2015, 198–200, 2017d; Gelber and McNamara 2016; Gelber 2017, 2019). However, we dispute the implied suggestion that defending hate speech laws on the grounds,

for example, that such speech can have silencing effects is akin to defending laws against alien abduction, the insinuation being that neither type of defence is supported by any hard evidence whatsoever. Instead, we point to not only a large body of evidence from legal cases and anecdotal evidence but also a growing number of social scientific studies involving interviews or questionnaires. Together this evidence points to many kinds of harmful consequences of hate speech, measured in terms of health, law and order, security, social cohesion, discrimination, oppression, subordination and access to public discourse and democratic decisionmaking (see Matsuda 1989b; Lawrence 1990, 1992; Matsuda et al. 1993; Nielsen 2002; Tsesis 2002, 2017; Delgado and Stefancic 2004; Parekh 2005–6, 2006, 2012; Langton 2012; Langton et al. 2012; Brown 2015, chs. 3 and 7, 2017d; Gelber and McNamara 2016; Gelber 2017, 2019).

All of this being said, we do not seek to make any sort of simplistic inference from the claim that hate speech can be harmful to the claim that it should therefore be prohibited (see also Boonin 2011, ch. 7). Rather, the fact that it is harmful simply warrants having a conversation about legal measures. It provides a sort of prima facie or narrow warrant, not an all things considered or overall warrant (see Brown 2015, ch. 1). After all, '[t]he harm principle might be a necessary, but it is not a sufficient reason for censorship' (van Mill 2017). To argue successfully that hate speech should be prohibited, all things considered, requires showing that a battery of arguments against banning it are flawed. We shall do some of this in Chapters 5 and 6.

1.4 Hate speech as a multisite problem

We have suggested that paying attention to country context means recognising that it makes less sense to speak of 'the problem of hate speech' than 'the problems of hate speech'. This is true for another reason. It is that problems of hate speech present themselves in many different areas, spheres or sites. In this section we try to sketch or map out some of the main sites in which problems of hate speech locate themselves: individual, group, society, technology, law, political figures and international relations.

1.4.1 Hate speech as an individual, group and societal problem

Hate speech is, on one level, an individual problem in that many hate speakers act as individuals and will often target other individuals for abuse, vilification, denigration or humiliation, either in face to face encounters or online. It is also the case that some of the harms of hate speech are inflicted on individuals, albeit the harms are often inflicted in a cumulative manner. Short-term emotional distress or fear, medium-term trauma and longer-term psychological or even physiological effects including anxiety, feelings of insecurity and low self-esteem – all of these harms can be experienced by individuals, even if these individuals are targeted because they are members of groups (see Matsuda 1989b; Delgado 1982, 1991; Nielson 2002; Barendt 2005; Tsesis 2010; Brown 2015, 49–58, 71–5; Gelber and McNamara 2016). This is why civil litigation has for a long time and continues to be a potentially appropriate avenue for some victims of hate speech in seeking redress (see Delgado 1982, 1983; Matsuda et al. 1993; Delgado and Stefancic 2018; Brown 2018b; Heyman 2018).

But on another level hate speech is clearly a problem at the level of groups and social networks. Many hate speakers are connected into wider hate groups or social networks of hatred, again either in person or increasingly online. Moreover, what they have to say (the hate speech) is never really about individuals but instead about entire groups, classes

of persons or sections of society. Even when they single out individuals for racist or sexist abuse, for instance, the individual is generally a mere representative of a group or groups. Thus, when black female political figures such as Diane Abbot in the UK or Cecile Kyenge in Italy have been subjected to racist and misogynist abuse online, they are seen by hate speakers as the living embodiment or encapsulation of everything that is allegedly bad about the wider group or groups to which they belong. Hate speakers are characteristically no respecters of the separateness of persons or the inherent individuality of human beings.

Hate speech is also clearly a problem of and for society. It is a problem of society in that it often manifests itself at societal levels. As we shall explore in Chapter 2, in countries like Italy, for example, migrants and refugees, especially from Africa, face widespread exposure to hate speech across society, including in the media, in politics, in the workplace, on the street and on the Internet. But hate speech is also a problem for society in the sense that it can be tackled at the societal level. In countries like Italy there are also civil society organisations that are working alongside government departments and concerned politicians to empower victims, ordinary citizens, social commentators and businesses to speak back against hate speech wherever it occurs in society.

Moreover, hate speech can do damage at the group and societal levels. For example, hate speech can pose a real danger to public order and security, it can reduce social harmony and community cohesion, it can play its part in structural and societal oppression or subordination, it can hinder intercultural dialogue, it can undermine public assurances that persons are members of society in good standing and it can call into question the political legitimacy of (systems of) rules and laws (see Tsesis 2002, 2017; Delgado and Stefancic 2004; Parekh 2005–6, 2006; McGowan 2012; Langton et al. 2012; Waldron 2012a, 2012b, 2017; Brown 2015, 2017d, 2018b).

Of course, how much one thinks hate speech is a problem of society depends on what one thinks about society itself. Those who see society – or their idealised vision of society or the nation – as being at a crisis point or under existential threat in some way due to immigration policies, for example, may not think it is such a great 'problem' if so-called hate speakers make derogatory remarks about immigrants. 'To make an omelette you have to break some eggs', they may say. But by the same token, the use of negative stereotypes, group defamation and false rumours about immigrants, especially by political figures, can contribute to or fuel a public perception of national crisis. By contrast, for those people who see the greatest problem faced by society as the occurrence of division itself, then hate speech is the disease, not the cure. Plus society itself is constantly evolving and changing, not least because of immigration. Patterns of immigration and changing demographics can bring about shifts in public attitudes about what counts as hate speech and whether society in general believes hate speech is morally unacceptable. We shall return to consider some of these changes in Chapter 2 and explore future changes in Chapter 9.

But all the while politics is never far from the picture. Hate speech at the societal level can be a problem for the operation of democratic politics. It can intimidate and inhibit certain sections of society from participating in public discourse and the formation of public opinion upon which democratic decisionmaking is based. And on controversial matters like immigration it can be hard for elected representatives and governments to identify genuine societal problems and formulate rational and reasonable policy responses to them amid the deafening noise of racist, xenophobic and anti-immigrant hate speech. In short, societal level hate speech will have an impact on the sort of politics we get, but the sort of politics we have will also over time shape the sort of society we end up with.

1.4.2 Hate speech as a technological problem

Hate speech is also a technological problem. Hate speakers have always used the latest technologies to spread their messages to as many people as possible, as cheaply as possible and as anonymously as possible. Printed leaflets, mail shots, automated telephone messages – these were just some of the technologies used by white supremacists and anti-Semites in the twentieth century. Today the Internet is home to massive quantities of hate content, with hate sites embedding links to other hate sites to create massive networks or systems of cyberhate. One effect of these systems has been to facilitate what Cass Sunstein calls 'cybercascades', where individuals are influenced to believe certain false rumours as a result of coming into contact with large numbers of like-minded believers online. Cascades can exist when individuals possess one or more of the following tendencies: to rely on information provided by other people because they lack access to it themselves, to substitute the judgment of speakers they esteem for their own personal judgment, to want or even emotionally need the good opinion of other people wherein adopting and spreading other people's beliefs is one way to endear oneself to members of an ingroup community (Sunstein 2007, 83–90).

In the past hate groups and independent hate speakers have been drawn to discussion board websites like 4chan, 8chan and Gab, but in order to add video content today they are increasingly posting the videos on YouTube and linking them to their posts on these other websites (Gantt Shafer 2017). Indeed, mainstream Internet social messaging and social networking platforms and websites like Twitter, Instagram and Facebook have become sites of vast quantities of hate speech content (Cohen-Almagor 2015), including published by political figures.

Of course, these Internet companies typically have in place community standards or terms of use which prohibit hate speech content (see Brown 2018a). In the UK, for example, Twitter, Instagram and Facebook have all taken decisions to remove accounts belonging to the EDL leader Tommy Robinson, citing contraventions of their community standards on hate speech (see Hern and Waterson 2019). (This was before Twitter adopted its new policy in June 2019 of no longer removing Tweets by political figures that violate its 'Hateful conduct policy' but instead simply requiring readers to view a generic warning label before clicking through to the relevant Tweets.). These moves were significant because in the past the EDL has relied heavily on social media to communicate internally, to get its message out and promote its causes externally, and to organise its events, including street demonstrations (Allen 2011). Nevertheless, still much of the work in implementing these community standards is ultimately done by human moderators. And this means that moderation is always retrospective and is very often slow and imperfect. It is also conducted in secret, meaning there is no external oversight of individual decisions (see also Brown 2018a). This is, of course, something that Facebook has recently sought to address through its plans to create an independent 'oversight board'.

Moreover, the human moderation of massive quantities of potential cyberhate invariably means some degree of infrastructure, such as native language speakers, local offices, training and management of moderators and so on. In countries like the UK, Facebook has built up the infrastructure needed to at least attempt to properly administer and enforce its community standards. However, other countries around the world where Facebook and other social media companies operate have seen a sharp growth in users, combined with high volumes of potential hate speech content. This has meant that Internet companies have failed to keep pace with the rate of cyberhate on their websites and platforms. So, for example, in her August 2018 Report on the Situation of Human Rights in Myanmar, the Office of the

High Commissioner for Human Rights (OHCHR) special rapporteur, Yanghee Lee, paid particular attention to the fact that Facebook is 'used extensively by much of the population in Myanmar and that, for many people, Facebook is the main way of using the Internet'.[19] Moreover, she expresses alarm at 'consistently high levels of hate speech, especially targeting religious minorities [i.e. Rohingya Muslims]'.[20] And she calls on Facebook to 'commit more resources to combating content that violates its own standards, in particular posts that incite hostility, discrimination or violence'.[21] Speaking in Geneva in March 2018 about her interim findings, Lee was even bolder in laying blame at the door of both Facebook users and Facebook itself for the climate of hostility in which mass atrocities against Rohingya Muslims have occurred.

> We know that the ultra-nationalist Buddhists have their own Facebooks and are really inciting a lot of violence and a lot of hatred against the Rohingya or other ethnic minorities. [. . .] I'm afraid that Facebook has now turned into a beast, and not what it originally intended.[22]

A related issue is the growing relationship between online hate speech and automated or artificial intelligence (AI) systems. Some scholars have pointed to how the alt-right has exploited automated systems or algorithms to promote hate content in search engine results (Daniels 2018). Whilst the challenges in using such AI systems to combat online hate speech are formidable, we have nevertheless seen Internet companies like Facebook invest large sums of money in trying to develop algorithms that can identity online hate speech, through learning as well as basic instructions, and then automatically remove this content or else flag it up for assessment by human moderators. In the future Facebook bots might even automatically generate and publish content which is intended to counter hate speech. But these AI developments in turn open up the possibility that hate speakers could themselves also use Internet bots to create and post online hate speech automatically. One undesirable upshot is human beings potentially being exposed to massive quantities of hate speech, far more than hate groups could manually produce. Plus, the Internet already decreases the chances of people actually meeting face to face to express their views, to see the whites of the other person's eyes, to recognise the feelings of pain and fear on the other person's face and to feel some empathy or even sympathy for the other person. If the content is created by hate chatbots, there is not even a human to see the suffering.

Aside from the issue of increased human exposure to hate speech and decreased human exposure to the effects of hate speech, another perverse upshot is that the hate chatbots might not be sophisticated enough to know if the content they create is being read by human beings or simply by other chatbots. And the anti-hate chatbots might not be sophisticated enough to know if the original content they are counter-speaking against was created by human beings or by other chatbots. So picture this scene: a webpage, online forum or set of social media posts and responses containing a pile of hate speech content and also a heap of counter hate speech content, none of which was created by an actual human being. Sitting behind this electronic facade of fake content would sit two algorithm developers, perhaps one in an anonymous city apartment block and another in a glass and steel office building in Silicon Valley. What is disturbing about this sci-fi scenario? When the conversation itself is automated, and does not involve an actual human being on either side, it takes away the opportunity for real people to truly learn something they did not know already through the medium of Internet communication. Part of being human is the ability to make the kind of *volte face* that even the cleverest bots can only dream of.

1.4.3 Hate speech as a legal problem

As well as being a social and technological problem, hate speech is also a legal problem, or a series of legal conundrums. One is simply whether or not to take legal measures to combat hate speech, that is, to make it unlawful under criminal, civil, human rights or anti-discrimination law. This includes assessments as to the efficacy of hate speech laws (Brown 2015, 242–51). Naturally part of the assessment is backward-looking and concentrates on infamous periods of racial hatred, such as the era of the Weimar Republic in Germany, and asks whether hate speech was rife in spite of hate speech laws because hate speech laws were not enforced in the protection of Jews and against anti-Semites or simply because laws restricting the use of hate speech, such as existed, were too underdeveloped or else developed too late to be capable of stopping what was taking place (see Doskow and Jacoby 1940; Riesman 1942, 728–9; Niewyk 1975, 112; Borovoy 1988, 50; Heinze 2016, 136–7). A related but often overlooked problem is how to measure the relative restrictiveness of legal measures compared to a myriad other extra-legal measures governmental authorities might also use to combat hate speech, such as education, empowering counter-speech or engaging in official counter-speech (see Brown 2015, 251–63).

Moreover, online hate speech raises vastly complex issues concerning not merely efficacy but also legal jurisdiction and legal responsibility for online hate speech content. One pressing issue is whether Internet companies merely provide mediums for online communication or are themselves publishers or editors of content and therefore can and should be held responsible for illegal content published on their websites and platforms (see Cohen-Almagor 2015; Brown 2018a; Scott and Delcker 2018).

A second conundrum is how to draft good hate speech laws (assuming this is possible). There are many technical challenges in drafting incitement to hatred laws, for example. This includes: how to specify the sorts of language that are relevant to stirring up hatred in a way that does not make the offences overbroad, such as whether to specify the offences in terms of threatening or abusive words or behaviour or something else; whether to base offences solely on intention or also on the likelihood of hatred being stirred up; and on what criteria to determine the proper scope of the offences in terms of the range of protected characteristic to be covered (see also Brown 2008, 2015, 26–8, 2016a, 2017c, 2017d, 607; Waldron 2017, 700–4).

A third conundrum is how to ensure that hate speech laws are interpreted and enforced in ways that serve their core functions or intended purposes. What that purpose is will depend not merely on the kind of hate speech law at stake but also on the country context. Thus, in Chapters 2 and 3 we shall try to explain how the particular historical, political, legal, economic, demographic, social and cultural circumstances of given countries plays into how the problem of hate speech is understood and, crucially, how legislators, judges and law-enforcement organisations have tended to interpret and enforce hate speech laws in those countries.

No doubt critics of hate speech laws, domestic or international, would argue that the real purpose is simply enabling a dominant liberal ideology, or even a dominant Western perspective, to silence dissent of alternative ideologies and regional worldviews. This may or may not be correct, but at least ostensibly legislators, judges and intergovernmental organisations 'make' hate speech laws for reasons they claim reflect legitimate state interests and ultimately serve the public interest. Thus, in countries which have Holocaust denial laws, the interest invoked is normally that of stopping the Holocaust or events like it from ever happening again. The deterrence need not be limited to the threat of prison sentences. It

can also be reputational. Some people may not want the reputational damage that comes from being engaged in or being perceived to be engaged in illegal activity (see also Kübler 1998, 361–2).

Yet it would be wrong to assume that the interests served by hate speech laws can always be expressed exclusively in terms of the state interest in deterring hate speech and preventing bad things from happening because of hate speech. Sometimes the purpose is more symbolic or even educative: a matter of sending a signal about what government thinks about hate speech (e.g. that it is abhorrent) and how it wants people to reconsider whether engaging in hate speech is acceptable behaviour (i.e. teaching people to consider the effects of hate speech on others) (see Brown 2015, 248–51). In addition, the purpose of extending the scope of existing hate speech laws, for example, can sometimes be to serve a norm – for example, the norm of parity of protection for different groups. In Chapter 3 we shall argue that this was one of the main purposes behind the Racial and Religious Hatred Act 2006 which created new stirring up religious hatred offences in England and Wales.

In short, it is crucial to move beyond generalities both about the meaning of the term 'hate speech' and the purpose of hate speech laws, to consider the particularities of given countries including the specific historical events and the wider patterns of social and political change which have significantly shaped how the problem of hate speech and the purpose of legal responses to it are understood therein.

1.4.4 Hate speech as a problem of, and not simply for, political figures

Another site of the problem of hate speech is the rhetoric of political figures. We shall use the term 'hyperpolitical hate speech' to describe hate speech when it is used by leaders of political parties or movements, political candidates, elected politicians and also unelected ministers and other public officials serving in government. In India, some Indian journalists call this 'VIP hate speech' (Jaiswal 2018). What makes it hyperpolitical and not merely political has to do with the status of the speaker as a political figure rather than the content of the speech. Now, clearly sometimes political figures use hate speech in addressing political issues or in articulating or representing political viewpoints. But not always or necessarily. Similarly, sometimes political figures can be the targets of verbal abuse or other sorts of vilification because of their politics or simply for being politicians. And it is a matter of debate whether speech about or targeted at people because of their political beliefs, affiliations or activities can ever be hate speech. We shall discuss that possibility in Chapter 5 [5.5.3]. But to reiterate, what differentiates hyperpolitical hate speech from other forms of hate speech is the speaker and not the content or the target of the speech.

For all the debate about what the term 'hate speech' means, that politicians and public officials sometimes use hate speech is virtually undeniable. Even INGOs like Human Rights Watch that have been critical of states for abusing or misusing hate speech laws to suppress political dissent (see Human Rights Watch 1992a, 1992b) have at the same time also highlighted the problem of political figures using hate speech against minorities or political enemies, including in countries as diverse as Saudi Arabia and the Côte d'Ivoire (see Human Rights Watch 2005, 2017). Several intergovernmental human rights organisations have also pointed to the role of political figures in using hate speech and contributing to a climate of hatred with their speech. For example, the CERD has drawn attention to 'the role of politicians and other public opinion-formers in contributing to the creation of a negative climate towards groups' through the use of racist hate propaganda.[23] Similarly, ECRI's executive body publishes periodic reports on member states of the Council of

Europe. Drawing on this evidence, the ECRI Commission has said this about the role of political figures.

> Although there have certainly been instances noted of political parties and other groups and organisations cultivating and disseminating racist, xenophobic and neo-Nazi ideas, the use of hate speech has not been limited to ones that are extremist and outside the mainstream. Thus, the employment of a rude tone in many parliaments and by state officials has been found to contribute to a public discourse that is increasingly offensive and intolerant. Such discourse has been exacerbated by some high-level politicians not being inhibited from using hate speech in their pronouncements. Furthermore, attempts by public figures to justify the existence of prejudice and intolerance regarding particular groups, which only tends to perpetuate and increase hostility towards them, have also been noted.[24]

Moreover, as we shall discuss in Chapters 2, 7 and 8, in some countries human rights commissions and equality courts have judged political figures to have infringed hate speech prohibitions contained in equality and anti-discrimination legislation. One well-known example is the South African case, *Afri-Forum and Another v. Malema and Others*,[25] involving the then leader of the ANC Youth League, Julius Malema, and the singing of the ANC struggle song 'Dubulu iBhunu' (Shoot the Boer/Farmer) at political rallies.

We shall also discuss several other cases involving political figures in other parts of the world. In Europe, for example, the ECtHR has recognised the special importance of the right to freedom of expression of elected representatives but has nevertheless upheld convictions under domestic hate speech laws precisely because of the influence elected representatives have on public discourse. Take *Féret v. Belgium*[26] involving the chairman of the Belgian political party Front National, Daniel Féret. During the late 1990s and early 2000s Féret was the editor-in-chief of the party's publications and a member of the Belgian House of Representatives. In a period of election campaigning between July 1999 and October 2001 Féret was responsible for the publication and distribution of leaflets which vilified immigrant communities as criminally minded and conspiring to exploit the Belgian benefits system. Despite his status as an elected representative at the time of the remarks, in 2006 he was convicted by Belgian courts for publicly inciting hatred, discrimination and segregation under Art. 444 of the Penal Code of Belgium and Art. 20 of the Law of 30 July 1981. Féret brought a complaint before the ECtHR alleging that his conviction was a violation of his human right to freedom of expression enshrined in Art. 10(1) of the European Convention on Human Rights (ECHR). According to this Article:

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.

However, the Court disagreed with his argument and affirmed that the relevant hate speech laws were 'necessary in a democratic society' under Art. 10(2) of the ECHR. According to this Article:

> The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial

> integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

The clear message from this and several other ECtHR cases is that in the context of worsening community relations, politicians have a moral obligation to avoid hate speech that fuels xenophobia and racism, for example, and risks intensifying the situation. That being said, in another set of cases, *Erbakan v. Turkey*,[27] *Perinçek v. Switzerland*[28] and *Stomakhin v. Russia*,[29] the ECtHR has decided that domestic convictions of political figures for hate speech offences had violated their right to freedom of speech under Art. 10(1) of the ECHR due to the particular circumstances of the cases allied to the way laws had been drafted. We shall return to these nuances in Chapter 8 [8.4.8].

There are innumerable reasons why political figures might engage in hate speech (see Brown 2019b), including the prosaic reason that sometimes hate speech directly reflects the politician's own bigotry and the more strategic reason that it can involve playing to the base. Whatever the reason, it would be naïve to think of political figures as lapsing into the use of hate speech accidentally, although this may happen. A recent academic study of the 2011 and 2015 general elections in Nigeria, for example, concluded that the use of hate speech by political elites during these elections was elevated to the status of a political campaign strategy (Ezeibe and Ikeanyibe 2017, 66). In Chapter 2 we shall compare the nature, extent and significance of hyperpolitical hate speech in a range of different countries, including Hungary, Italy, Turkey, South Africa, Nigeria, Kenya, the UK and, of course, the US.

Furthermore, in Chapter 7 we ask the normative question whether hyperpolitical hate speech is morally problematic in any distinctive ways compared to ordinary hate speech, and whether political figures have any special moral duties to refrain from the use of hate speech over and above general moral duties ordinary people may have. In particular, we shall investigate whether hyperpolitical hate speech, like other political phenomena such as democracy itself, has particular characteristics relating to power, authority, influence, deliberation, representation, leadership, conflict, personal ambition, bureaucracy and legitimacy that feed into the normative picture. Reflecting on these characteristics the issue becomes whether hyperpolitical hate speech involves distinctive wrongdoing or morally bad consequences, exhibits a unique concatenation of general wrongdoings or morally bad consequences, exemplifies higher degrees of wrongdoing or moral badness or else simply implicates a singularly complex combination of rights and wrongs and of morally good and bad consequences.

Another part of the story, to be discussed in Chapter 8 [8.2.4], is whether there can be circumstances in which it is correct for political figures to do things (dirty hands) or to say things (dirty words, if you will) that would otherwise be morally wrong. Consider once again Trump's rhetoric during the 2016 presidential election campaign.

> When Mexico sends its people, they're not sending the best. They're not sending you, they're sending people that have lots of problems and they're bringing those problems. They're bringing drugs, they're bringing crime. They're rapists and some, I assume, are good people, but I speak to border guards and they're telling us what we're getting.[30]
>
> Hey, I watched, when the World Trade Center came tumbling down, and I watched in Jersey City, New Jersey, where thousands and thousands of people were cheering as that building was coming down. Thousands of people were cheering.[31]

Just imagine for the sake of argument that during the 2016 campaign Trump actually knew things about the threats facing America that the American people did not know. And suppose for the sake of argument that Trump sincerely believed that a wall had to be built and a travel ban had to be imposed on people coming from Muslim countries in order to keep America safe. And also suppose that using negative stereotypes against Mexican immigrants and spreading false rumours about Muslims actually was, under the circumstances, the most persuasive way of mobilising public support, thus improving Trump's chances of becoming president and doing what needed to be done. This may not be pretty, but it is real politics (some people might argue).

Of course, there are a lot of highly questionable assumptions here. In the case of Trump there is a fine line – sometimes invisible – between sincerely held beliefs about what is needed to keep America safe and what must be said to put him in a position to achieve this, on the one hand, and political calculations about what he needs to say in order to maximise his chances of winning elections, period, on the other hand. What is more, even if an argument could be made about a national emergency which somehow requires the use of hate speech against Mexicans and Muslims, what national emergency exists that could possibly warrant the use of such language in relation to women or disabled people (see also Brown 2019b)?

So we might be happy to play along with the philosophical fantasy of exceptional circumstances or emergencies situations in which a high moral threshold needed to warrant political figures in using hate speech has been met. But in reality this high moral threshold is rarely met. Narratives presented as national emergencies that could potentially take precedence over moral obligations to refrain from the use of hate speech are often not national emergencies at all but are simply part of the rhetorical architecture of the hate speech itself, part of the web of negative stereotypes and false rumours.

We take inspiration from Bruce L. Payne's astute moral evaluation of the behaviour of Spiro Agnew, the former vice president of the United States:

> At his worst Agnew appealed to fear and anger, and skirted the edge of racist demagoguery. These are common tactics of narrowly self-interested opportunism: emotional appeals for the most part are too briefly effective to support a long-term program or an enduring party organization. Individuals intent on winning can regularly benefit from such a course, though only at the price of unacceptable risks to the stability and decency of the public order.
>
> (Payne 1981, 184)

There are some irresistible parallels to be drawn between Agnew and presidential candidate and now President Trump.[32] And, like Payne, in Chapter 7 we shall argue that political figures such as Trump, but also those from the many countries we discuss in the book, have a moral duty and professional responsibility to resist the temptation to engage in hate speech in pursuit of personal political success. In reality this obligation is not in spite of the good of the nation but because of it. It is a matter of not contributing to a climate of hatred in which acts of discrimination and violence are more probable (see Chapter 7 [7.3]) and not legitimising and normalising ordinary hate speech (see Chapter 7 [7.8]).

Moreover, in Chapter 8 we shall challenge the assumption that the speech of political figures is so special, democratically speaking, that it is either impossible or unconscionable to hold them accountable under hate speech laws along with ordinary people. We argue that

it is precisely *because* political figures have special power, authority and influence that they should not enjoy immunity from hate speech laws. And if it is absolutely necessary in a democratic society for there to be parliamentary privilege, then at the very least parliamentarians should be subject to quasi-legal sanction procedures relating to their use of hate speech. The same goes for people seeking elected office. And so we recommend a strengthening of the enforcement and, where necessary, changes to the content of party, parliamentary and election codes of conduct in relation to hate speech.

1.4.5 Hate speech as an international problem

Like many problems facing the world today, hate speech is not simply a domestic problem but also an international problem, and in two important senses. First, the globalisation of hate speech, especially through digital media, means that hate speech does not respect state boundaries, if it ever did. This speaks to the technological nature of the problem of hate speech discussed prior. A particular state could unilaterally adopt a raft of measures to combat hate speech, including criminalising certain forms of hate speech. But that would not prevent hate speech pouring into the country from sources in neighbouring countries where hate speech is not unlawful, including through the radio, television, newspapers and increasingly over the Internet. A state could, as Germany has done and many other states, including France and the UK, are considering, introduce Internet laws that hold companies responsible for failing to take down 'obviously' illegal hate speech content within a specified short period of time. Such laws are an attempt at the extraterritorial regulation of Internet companies. But if only a handful of countries adopt such measures, this leaves companies operating in countries without such laws free to carry on failing to remove illegal hate speech content, the wider effects of which can still cross borders. Nor do such laws punish or deter the third parties who actually create the content, who might live in countries with inadequate or unenforced hate speech laws. This leaves these individuals free to carry on posting new content under new accounts, in a game of cat and mouse with authorities in foreign countries. States ostensibly join international conventions or treaties relating to hate speech partly because of the communicative interconnectedness of states and in the hopes of mitigating the effects of the globalisation of hate speech. We shall speak more about this in Chapter 4.

Second, hate speech and the drafting and implementation of international hate speech instruments can often be caught up in wider geopolitical disputes across multiple countries and even entire regions. Sometimes hate speech crops up in the context of historic conflicts, tensions or disputes, including as a legacy of imperial attitudes. For example, in December 2014 the South Korean parliament adopted a resolution requesting that Japan 'crack down on "hate speech" demonstrations targeting Korean residents' in Japan (Koontz 2017, 348). Importantly, when the South Korean parliament passed this resolution it invoked the work of the CERD as a justification. The CERD undertakes periodic reviews looking at progress (or lack thereof) made by States Parties in implementing the ICERD. One of the areas the CERD considers is whether states have introduced domestic hate speech laws in accordance with commitments agreed by States Parties under Art. 4 of the ICERD.

> States Parties condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and

> undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in article 5 of this Convention, inter alia:
>
> (a) Shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof;
>
> (b) Shall declare illegal and prohibit organizations, and also organized and all other propaganda activities, which promote and incite racial discrimination, and shall recognize participation in such organizations or activities as an offence punishable by law;
>
> (c) Shall not permit public authorities or public institutions, national or local, to promote or incite racial discrimination.[33]

In its 2014 report on Japan, the CERD called on Japan to review and consider withdrawing its reservation to Art. 4 of the ICERD and recommended that Japan 'take appropriate steps to revise its legislation, in particular its Penal Code, in order to give effect to the provisions of [Art.] 4' (CERD 2014, para. 10). The CERD offered as grounds for these recommendations the following concerns.

> The Committee is concerned about reports of the spread of hate speech including incitement to imminent violence in the State party by right-wing movements or groups which organize racist demonstrations and rallies against foreigners and minorities, in particular Koreans. The Committee is also concerned by reports of statements made by public officials and politicians amounting to hate speech and incitement to hatred. The Committee is further concerned by the propagation of hate speech and incitement to racist violence and hatred during rallies and in the media, including the Internet. Furthermore, the Committee is concerned that such acts are not always properly investigated and prosecuted by the State party.
>
> (Art. 4)

On other occasions internationally significant hate speech occurs in third party states, meaning that citizens of third party states engage in hate speech to protest against the situation of peoples involved in historic conflicts, tensions or disputes elsewhere in the world. This has often been the case with the Israel-Palestine conflict. Consider *Willem v. France*.[34] In 2002 Jean-Claude Willem was a member of the French Communist Party and mayor of the French municipality of Seclin. During a town hall meeting he announced his intention to ask the catering services for the municipal council to boycott Israeli products, especially fruit juice, as a political protest against what he believed were repressive policies against Palestinians being pursued by the government of the State of Israel and, in particular, the prime minister, Arial Sharon. The French Court of Cassation upheld his conviction for provocation of discrimination on national, racial or religious grounds under Arts. 23 and 24 of the Law on the Freedom of the Press of 29 July 1881 (as amended[35]). Willem lodged a complaint against this decision with the ECtHR in 2005. He argued that his call to boycott Israeli products was part of a political debate concerning the Israeli-Palestinian conflict and was a matter in the general interest and, therefore, that his conviction was a violation of his human right to freedom of expression under Art. 10(1) of the ECHR. However, the

ECtHR held that the conviction was proportionate to the legitimate aim pursued under Art. 10(2) of the ECHR. Specifically, the ECtHR agreed with the French prosecutors and Court of Cassation in considering that Willem's comments in the context in which he made them as a mayor were not merely an expression of a political opinion but a provocation to discrimination addressed to catering services working for the municipal council.[36]

We shall discuss the implications of these sorts of cases and issues for questions of International relations and international law in Chapter 4. We consider several important questions. Why do some states and not others agree, sign, ratify and comply with international hate speech instruments? Why does any state accept the jurisdiction of international human rights courts or intergovernmental organisations in hearing complaints against it relating to domestic hate speech laws? What role do international norms on hate speech play in the evolution of international hate speech instruments? What part do norm entrepreneurs have in developing these norms in the first place? In answering these questions we shall draw on several approaches to international relations theory: realism, institutionalism, constructivism and critical approaches. We defend a pluralistic account drawing on elements of each approach.

In Chapter 4 we also explore in depth the evolution of US foreign policy in relation to international hate speech instruments. American statespersons have often made the argument on the international stage that domestic hate speech laws can be dangerous if they provide a cloak of respectability to authoritarian or illiberal regimes in crushing political dissent they portray as 'hate speech'. At the same time the US has been reluctant to acknowledge the rational limits of that warning: that the danger of the abuse or misuse of hate speech laws is not always a clear and present danger. As a result, the US has ratified international hate speech instruments but entered reservations which mean they need do nothing to implement the requirements of those treaties. We argue that, ironically, this is a case of the US using its ratification of the treaties to give itself a cloak of respectability in the eyes of the rest of the world. We also consider how the US position might change in the future, perhaps as a reaction to the disruptive diplomacy of Donald Trump, which is characterised by his not merely declining to properly engage with hate speech treaties but actually using hate speech against other countries ('shithole countries').

In Chapter 4 we also look at how diplomatic criticism is one of the currencies of international disputes over hate speech. This raises several further issues. For example, why do states engage in diplomatic criticism? How, if at all, is diplomatic criticism effective? In answering these questions we once again draw on the aforementioned approaches, but also some English School approaches. We highlight the importance of identity and reputation, as well as the notion that states may simply choose to be good international citizens. Following on from this, we consider whether states have a moral obligation to engage in diplomatic criticism of other states who transgress against the international norm on hate speech. We argue that this obligation to be norm watchdogs itself derives from a more fundamental or general moral obligation, namely, the moral obligation to support just international norms.

In addition, we consider what obstacles can stand in the way of diplomatic criticism having its desired effect. One such obstacle we examine is when states making the criticism face counter-accusations of hypocrisy. Suppose the Italian economy suffers a catastrophic event that sends inflation and unemployment through the roof. Many Italians migrate to Germany in search of work and cheaper cost of living. Italophobia becomes prevalent in some parts of Germany. Far-right German politicians negatively stereotype Italians as prone to corruption, as being unable to manage their finances, as putting la dolce vita above hard work and so on. Some German politicians stir up hatred against Italians by falsely accusing them of

raping German women. Some incite Germans to block Italians from purchasing or renting apartments in certain municipalities. Suppose German courts find some of these politicians guilty of hate speech offences under s. 130 of the German Criminal Code. But then at the time of sentencing judges use their discretion to suspend the sentences in all cases citing the fact that the politicians had already made public apologies and were performing civic roles as elected representatives. Now suppose the state of Italy takes the step of publicly criticising the state of Germany for failing to do enough to combat hate speech against Italians living in Germany. However, the state of Germany for its part ridicules Italy for its hypocrisy. It points to the not uncommon practice among Italian courts of handing down suspended sentences for politicians found guilty of hate speech offences. It also draws attention to the fact that because of the suspended sentences these politicians have been able to carry on with their political careers, win elections and hold elected offices, as though nothing had happened. We shall discuss just such cases in Chapter 2 [2.13]. What obstacles does this sort of hypocrisy pose to the ethical standing and wider persuasive force of diplomatic criticism? In Chapter 4 we shall argue that charges of hypocrisy [4.5.4], as well as accusations of postcolonialism and neo-colonialism [4.5.5], although important, are not insurmountable obstacles to the ethical standing and persuasive force of diplomatic criticism insofar as other states can separate the criticiser from the criticism.

1.5 Preliminaries

Before we begin trying to make good on the aforementioned promised lines of inquiry and substantive arguments to come, we first need to say something of the different methodologies we employ and of the canon of academic research, and wider academics debates, into which our investigation fits.

1.5.1 Methodological framework

At the start of the chapter we proposed that there are several reasons to be optimistic that genuine progress can be made, and is being made, on the issue of hate speech and hate speech laws: signs that political communities and the international community as a whole are already moving beyond seemingly entrenched positions and intractable disputes on this issue and are capable of moving even further. In the book not only do we seek to explain and provide evidence of these developments, but we also try to justify them, that is, to show why they are, mostly, albeit with qualification, good, just and legitimate. This is a hybrid project, therefore, combing both descriptive and prescriptive elements.

In the descriptive part of the project we frequently point to the emergence of and adherence to norms in order to explain and understand the developments we are talking about. These norms are conventional understandings, practices and expectations among actual peoples, communities or organisations, including at both the domestic and international level, concerning standards of appropriate behaviour for actors with given roles or identities (see also Finnemore and Sikkink 1998, 891). In other words, this sort of normative scholarship is normative in the technical sense that it explains developments in terms of conventional standards or norms.

In the prescriptive part of the project we make normative arguments from within the Anglo-American analytical tradition of philosophy, which aims to analyse normative values, principles and obligations and to identify moral truths about what sorts of behaviours and what sort of institutions and laws are good, just or legitimate. This in turn leads

to recommendations about how behaviours, institutions and laws should be reformed to become better, more just or more legitimate (see Pettit 1993; McDermott 2008). In particular, we employ insights and techniques from normative jurisprudence, political philosophy and applied ethics to theorise the values, principles and obligations at stake in debates over hate speech and then apply these in evaluating and recommending reforms to behaviour, institutions and laws surrounding hate speech (see also Heyman 2008; Waldron 2012a; Brown 2015, 2018b, 2017d). However, the approach we take in this book tends not to be one that takes certain values, principles or obligations as given and then works through the implications for particular circumstances (see Cohen 2011). Rather, we often start with observations about particular institutions, norms, structures and practices, as well as the contexts in which they are located, and build normative principles that are bespoke for the appropriate regulation of those things (see Miller 2008).

Having identified these two parts to the project we also need to make the following caveat: it would be wrong to assume that there is always clear daylight between the sort of thinking that international relations theorists get up to when they study the emergence of international human rights norms, say, and the sort of normative argumentation characteristically done by normative legal and political philosophers and applied ethicists who turn their attention to questions – in both ideal and non-ideal theory (Simmons 2010) – of justice, legitimacy, human rights and so on. For example, the idea that conventional constructivism could be 'value-neutral' is belied by the fact that the relevant scholars invariably make choices, conscious or otherwise, about which norms matter and which are probably forces for change of the morally good variety. Indeed, critical constructivism would argue that moral critique is a key part of the constructivist toolkit alongside explanation and analysis: the point of critical constructivism is not merely to understand international relations but to make things morally better such as by unmasking inequalities of power, injustice and oppression and working towards their eradication, according to the moral vision not merely of the agents being studied but also of the critical constructivist (see Sinclair 2010). Similarly, Solidarist English School approaches to international relations are often quite explicit in the fact that they are laying claim to moral truths about good and bad, right and wrong, just and unjust (see Wheeler and Dunne 1998). In the words of Richard Price, 'normative theorizing is inescapably involved in making or challenging claims about possibilities of moral change in world politics, and thus ethics is central to practice and intellectual discourse in international relations, even as professionally it has not been accorded pride of place in the American academy of international relations, which has been dominated by explanatory agendas that have largely excluded normative theorizing as the terrain of "political theory/philosophy," "normative theory," or philosophy' (Price 2008, 193).

Having made this point about normative overlap, what should be relatively obvious is that a multidisciplinary approach is vital for tackling the sort of hybrid project we undertake in the book. The approach we take reflects our varied methodological backgrounds. It combines: normative jurisprudence and legal theory, analytical political philosophy and political theory, applied ethics, sociology and social theory, comparative law, critical legal studies, public policy theory, political science and area studies (Alexander Brown); and international relations theory, international law scholarship, international legal theory, comparative legal scholarship, critical legal studies, political science and area studies (Adriana Sinclair).

In chapter 2 we employ an interpretivist and contextualised approach to understanding the problem or problems of hate speech in different countries (see also Bayard de Volo 2016). This involves understanding the meaning and salience of the problem of hate speech in each country and identifying how this meaning and salience not only shapes policy and institutional responses to hate speech but is also shaped by each country's particular historical,

political, legal, economic, demographic, social and cultural circumstances, including different sensitivities, perspectives or standpoints, narratives and ideological frameworks. In terms of the book as a whole, the main focus is on the UK and the US, but these countries are far from being the sole focus. We appeal to examples from many countries in the course of the book. Thus, in order to provide background context for these examples, in Chapter 2 we explore in detail 12 country contexts: Nigeria, Kenya, South Africa, India, China, the US, Japan, the UK, Turkey, Germany, Hungary and Italy. In this chapter we make use of insights and methodologies in sociology and social theory, comparative political science, comparative legal scholarship and area studies.

In Chapters 3 and 5 we draw on several approaches and techniques from political science as well as some approaches in public policy theory. In assessing the evolution of public policy on hate speech we employ not only interpretivist approaches in looking for the meanings attached to the problem of hate speech by policymakers but also some more positivist approaches, most notably the method of diachronic analysis. This involves using a temporal or time-dependent perspective to understand the evolution of policymaking at a fine-grained level (see also Howlett and Rayner 2006). Specifically, in Chapter 3 we look at the temporal position of the Racial and Religious Hatred Act 2006 relative to facts about particular antecedent historical events and facts around when certain discourses were or were not prominent in the governing party in the UK. We then use this temporal sequence as a basis for judgments about which of several competing explanations as to the real function of the legislation is most credible (albeit without assuming anything as strong as path dependency). Explanations of the real function of the legislation that rest on or appeal to historical facts and discourses that actually occurred after the legislation was enacted will be considered less credible than those which explain the function of the legislation in terms of historical facts and discourses that occurred prior to or at the time of the birth of the legislation. Likewise, in parts of Chapter 5 we also employ temporal perspectives to challenge the slippery slope objection to hate speech laws. We look at various countries in which the actual sequence of historical events does not support the idea that hate speech laws lead to massive censorship.

In Chapter 4 we begin by employing methods in international law scholarship to compare and contrast different international hate speech instruments in terms of content and in terms of the distinction between hard law and soft law. And we look at processes of convergence between different international hate speech instruments but also between international hate speech law and domestic hate speech law, including judicial globalisation. International legal scholarship has its own divide between approaches which seek to clarify what international law says and how it should be applied to particular cases or issues and normative approaches which offer a prescriptive judgment on the content of those laws. Are they the right laws? How should they be changed to make them more just or more legitimate (Diehl and Ku 2010)? We shall engage in both forms of international law scholarship in the course of the chapter. For example, in the conclusion of the chapter we shall argue that international hate speech instruments should be reformed to better reflect the broad scope of domestic hate speech law in terms of the range of protected characteristics covered.

Chapter 4 also attempts to understand why some states do and some states do not agree, sign, ratify and comply with international hate speech instruments, and it does so by drawing on a plurality of different approaches to international relations theory. For international relations theories the most pressing question is what effect international law has on state behaviour: is it a constraint? If so, how does it constrain behaviour? And why do states bother with international law at all? Like most questions asked by international relations theorists, questions about international law generally boil down to questions about power: 'what kind of

power does international law have?' (Byers 1999; Reus-Smit 2004). For many international legal scholars, by contrast, the power of international law is taken as a given.

More broadly, international relations theory is an attempt to understand international politics by providing generalised and overarching explanations for why states behave in certain ways and why relations between states but also between states and other international actors such as international organisations and INGOs take the forms they do. The different approaches to international relations theory we use in Chapter 4 also straddle an important divide in international relations scholarship between empiricist natural science methods like rationalist positivism and interpretivist methods like reflectivist post-positivism (see Hollis and Smith 1990). Rationalist, empiricist and scientist approaches will develop and test hypotheses, just like in the natural sciences, often using statistics to do so, and posit assumptions about states as rational actors. Realism and institutionalism fall broadly into this camp. Interpretivist approaches, however, do not believe that we can have objective knowledge of the world, partly because what each of us sees or interprets is coloured by who we are. As a consequence, the pseudo-scientific method of creating and testing hypotheses is not only unreliable but, worse still, purports to provide definitive explanations. For interpretivists, in order to say anything meaningful about the world, the researcher must reflect upon who he or she is, but also what it means to be human and how that shapes the world around us. Interpretivism is also typically committed to uncovering the operation of power in international politics. As such, critical approaches exemplify this sort of methodology. Constructivist approaches provide a bridge between this methodological divide, with some elements falling on the rationalist empiricist side and some on the interpretivist side.

In Chapter 4 we also consider the normative question of whether states have moral obligations to engage in diplomatic criticism against states that fail to comply with the international norm on hate speech. And we examine what it means for states to have or lack ethical standing to perform such criticism. Here we engage in the sorts of normative analysis and argumentation characteristically done by normative legal theorists, political philosophers and applied ethicists described prior. This includes analysis of the nature of moral obligations in the international domain, such as the distinction between primary and secondary obligations in relation to global justice (see Nagel 2005), and arguments about obstacles to ethical standing, such as hypocrisy (see Cohen 2013).

Chapter 6 challenges a range of standard objections to hate speech laws found in the literature. These standard objections typically appeal to practical considerations of various kinds, based on methods of comparative law, political science and public policy, but also to distinctively legal principles and doctrines, to important political values and institutions like democracy and to wider ethical considerations, of the variety found in normative jurisprudence, political philosophy, democratic theory and applied ethics (see Lively 1994; Strossen 2001, 2012, 2018; Weinstein 1999, 2001, 2012a, 2017a, 2017b; Baker 2009, 2012; Dworkin 2009, 2012; Heinze 2016). Our responses to those objections also draw on practical considerations based on methodologies in public policy and political science as well as a number of legal, political and moral values, including equality, freedom from oppression, human dignity, civic dignity and political legitimacy, for example (see Delgado and Yun 1994a, 1994b; Delgado and Stefancic 1996, 2004; Heyman 2008; Brown 2015, 2016a, 2017c, 2017d, 2018b).

Chapters 7 and 8 make the case, first, that it can be morally problematic for political figures to engage in hate speech, and, second, that it would be politically legitimate for governments and those in charge of political parties and parliamentary behaviour to take legal and quasi-legal measures against hyperpolitical hate speech, including limiting the scope of

parliamentary privilege to exclude hate speech laws, and changing and strengthening the content and enforcement of party, parliament and election codes of conduct (see also Hill 2007; Lakin 2013; Thompson 2018). Here we employ methodologies in normative jurisprudence and legal theory, analytical political philosophy and political theory, political ethics and comparative law.

1.5.2 The canon

In this subsection we briefly outline the canon of existing scholarly work to which this book simultaneously owes a debt, seeks to challenge and attempts to make an original contribution. First, there is literature on the concept of hate speech. Some of this work sets forth core elements or features of what an ideal legal concept of hate speech would look like, including the sorts of harmful or discriminatory speech acts it would make unlawful (see Matsuda 1989b; Brison 1998, 2013; Tsesis 2002; Delgado and Stefancic 2004; Cohen-Almagor 2009; Langton 2012; Langton et al. 2012; Gelber 2012, 2017, 2019). By contrast, Brown (2017a, 2017b) provides a total conceptual analysis that distinguishes between the legal concept and the ordinary concept, challenges the myth that hate speech is essentially about hatred and argues that hate speech is a family resemblances concept. Earlier in this chapter we explored one aspect of this analysis: the politics and politicisation of the ordinary concept hate speech.

Second, there is a growing literature on how the particular political, legal, social, economic and cultural circumstances of a country shape how putative problems of hate speech emerge and whether they are either accepted as problems to be tackled or downplayed as social overreactions or merely the necessary price of freedom of expression. In his groundbreaking studies Walker (1990, 1994) looks at the history of hate speech controversies in the US and makes important contributions to understanding the potentially surprising forces that opposed hate speech laws in the second half of the twentieth century – most notably civil rights organisations representing African Americans. In a similar sort of vein, Goldberg (2015) explores how in Germany during the 1890s a new vision of hate speech began to take shape when grassroots Jewish social movements began to utilise laws in order to oppose anti-Semites. Of note is also the work of Bleich (2011), which in fine-grained terms compares public policy approaches to hate speech in Europe and the US, placing hate speech laws into political and social contexts. In addition, the excellent work of Boromisza-Habashi (2013) and Molnar (2010, 2012) investigates public discourse on the issue of hate speech in Hungary and looks at how civil society, the parliament and the courts have each responded to the issue in that country. Analyses of specific countries can also be found in one of the great compendium volumes on hate speech laws edited by Hare and Weinstein (2009).

Pushing these envelopes further, in Chapter 2 we compare and contrast the issue of hate speech in 12 different countries – Nigeria, Kenya, South Africa, India, China, the US, Japan, the UK, Turkey, Germany, Hungary and Italy. We describe the role that disputes over hate speech and hate speech laws have played in national life, but we also attempt to explain how particular contextual circumstances have in turn shaped those disputes – including colonialism and post-colonialism, the fall of communism, large-scale immigration, changes in cultural attitudes and morals, the rise of personality politics and political popularism and wider regional conflicts and tensions. For example, Chapter 2 looks at countries like the US, Italy, Hungary, Turkey, Kenya and South Africa as case studies of where it is especially apparent that the personal ambitions and charisma of individual politicians has enabled them to appropriate and recast these disputes to serve their own political ends. Focusing on a

different type of context, Chapter 3 looks at the social and political forces that led to the introduction of the stirring up religious hatred offences in England and Wales. Pace Gelber (2018), we emphasise not concerns over home-grown terrorism but historic arguments about public order and parity of protection.

Third, many international legal scholars have already sought to compare and contrast the structure and content of different regional and international hate speech frameworks or instruments, typically from an analytical legal perspective (see Dore 1981; Defeis 1992; Mayer-Schönberger and Foster 1995; Delgado and Stefancic 2004, ch. 12; Bertoni and Rivera 2012; Thornberry 2016; Keane and Waughray 2017b; Dashtevski and Ilieva 2017; Alkiviadou 2018). In Chapter 4 we attempt to build on but also go beyond this existing literature and its black-letter approach to studying international law. Not merely do we look at a more comprehensive sample of regional and international hate speech instruments in order to highlight the importance of regional context, but we also seek to make some legal theoretic arguments about judicial globalisation and legal convergence.

Fourth, there is a small body of work looking at the international politics surrounding the negotiations on drafting and ratification of key international hate speech instruments, including the International Covenant on Civil and Political Rights (ICCPR) and the International Convention on the Elimination of Racial Discrimination (ICERD) (see Meron 1985; Kübler 1998; Mchangama 2011). However, some of this work tends toward more historical and descriptive approaches, while other contributors tend to concentrate on realist approaches to international relations theory. Chapter 4 examines why some states and not others agree, sign, ratify and comply with international hate speech instruments through the lenses of realist, institutionalist, constructivist and critical approaches to international relations theory. In doing so we seek to make a contribution not merely to our understanding of international hate speech instruments but also to some extent to the strengths and weaknesses of these approaches as explanatory tools.

Related to this, in Chapter 4 [4.5] we also explore the issue of diplomatic criticism and whether states have a moral obligation to engage in such criticism of states who fail to comply with international hate speech instruments. We begin by exploring this question descriptively and invoke once again the aforementioned approaches but also some English School approaches. But we then also look at this from the perspective of normative philosophy and, specifically, theories of moral obligation and theories of ethical standing. Based on this synthesis of different disciplinary insights and approaches, we seek to make two original contributions to scholarship on diplomatic criticism. First, we argue that states have a fundamental or general moral obligation to support just international norms and that this grounds two kinds of particular moral obligations: primary obligations to adhere to those norms and secondary obligation to engage in diplomatic criticism of other states who fail to adhere to those norms. Second, we highlight three problems that could potentially undermine the ethical standing and/or persuasive force of diplomatic criticism concerning hate speech: the problem of norm contestation, the problem of counter-accusations of hypocrisy and the problem of counter-accusations of postcolonialism and neocolonialism. We argue, however, that all three problems are not insurmountable.

Finally, there is a vast jurisprudential literature looking at the moral, legal, social and political pros and cons of regulating hate speech. This literature examines criminal hate speech laws, including group libel and incitement to hatred laws (see Parekh 2005–6; Hare and Weinstein 2009; Strossen 2001, 2012, 2018; Weinstein 2001, 2011, 2017a, 2017b; Herz and Molnar 2012; Waldron 2010, 2012a; Botha and Govindjee 2014; Brown 2015, 2017d; Heinze 2016). It also considers discriminatory harassment laws, including campus speech codes (see

Lawrence 1990; Lange 1990; Strossen 1990; Delgado 1991; Altman 1993; Delgado and Stefancic 1994; Delgado and Yun 1994a; Shiell 2009; Tsesis 2010, 2017). It explores discriminatory intimidation and provocation laws, including cross-burning statutes (see Lawrence 1992; Amar 1992; Greenawalt 1995; Shiffrin 1999; Hare 2003; Hartley 2004; Heyman 2008; Brown 2015). It investigates civil remedies for victims of hate speech (see Delgado 1982, 1983, 2018; Heins 1983; Matsuda et al. 1993; Delgado and Stefancic 2018; Brown 2018b; Heyman 2018). It also looks into informal hate speech procedures, where complaints can be brought to, and heard by, equality courts or human rights commissions (Gelber and McNamara 2014; Botha and Govindjee 2016). In addition, it is increasingly concerned with media and Internet laws, requiring Internet companies to remove illegal hate speech content (see Cohen-Almagor 2015; Brown 2018a; Scott and Delcker 2018).

This scholarship has also discussed at length certain types of situations connected with some instances of hate speech which potentially cut across all of these laws. Consider the literature on captive audiences to hate speech (see Matsuda 1989b, 2372–3; Lawrence 1990, 456; Battaglia 1991, 376; Strauss 1991, 89–103; Jones 1992, 4; Massey 1992, 177; Eberle 1994, 1211–12; Balkin 1999, 2312; Sadurski 1999, 186; Brink 2001, 135; Reichman 2007, 120; Heyman 2008, 165–6; Corbin 2009, 962–3; Delgado and Stefancic 2009, 362; Shiell 2009, 110–11; Weberman 2010; Berger Levinson 2013, 67; Brown 2017e).

This book does not seek to rehash all of these debates, but instead to highlight some areas that have been hitherto either misconceived or simply ignored. In terms of misconceived areas, in Chapters 5 and 6 we challenge a series of arguments that have been made against hate speech laws in the literature, including but not limited to the slippery slope argument (that hate speech laws lead to more laws and ultimately massive censorship) and the balkanization argument (that hate speech laws increase rather than decrease hostility because they encourage people to concentrate on identity differences even more). We argue that these and many other arguments against hate speech laws are both unsound (lacking in evidence) and invalid (draw conclusions that simply do not follow).

In relation to overlooked areas, we note a common feature of the problem of hate speech in nearly all of the country contexts considered in Chapter 2, namely, the use of hate speech by political figures. Although this problem has been highlighted repeatedly by some human rights INGOs and by intergovernmental human rights organisations such the CERD and ECRI, it has hitherto received very little serious academic attention, albeit with some exceptions (see Hill 2007; Lakin 2013; Thompson 2018). Chapters 7 and 8 attempt to put that right. Chapter 7 examines what it is about political figures in terms of their power, authority and influence that makes hyperpolitical hate speech uniquely or especially morally problematic, and argues that political figures have special moral duties to refrain from the use of hate speech. Following on from that, Chapter 8 puts forward and defends a range of new proposals for legal and quasi-legal measures to combat hyperpolitical hate speech, including limiting the scope of parliamentary privilege to exclude hate speech laws, and changing and strengthening the content and enforcement of party, parliamentary and election codes of conduct.

1.5.3 The wider field of academic literature

Whilst we focus on the politics of hate speech laws, we also keep in mind the wider context of academic debate to which many of the issues discussed in the book are connected. First, in Chapter 2 we explore the significance of hate speech and hate speech laws in local context including a discussion of how hate speech laws are enforced by public prosecutors and the courts. In many countries there is concern about politically or ideologically partisan courts,

and in some countries there are accusations of judicial capture by the government. This speaks to wider academic debates about the credibility of assumptions of judicial independence and neutrality, but also discussions on how to safeguard and promote these values (see Vanberg 2008; McHarg 2019).

Second, some of our examination of international relations theory and international hate speech instruments in Chapter 4 touches on broader empirical work on what sorts of states ratify human rights treaties. For example, research has shown than whilst authoritarian regimes are just as likely as liberal democracies to sign and ratify human rights treaties, ratification has been shown to have little effect on actual human rights practices in certain authoritarian regimes, unlike in democratic states (see Hafner-Burton et al. 2008; Hathaway 2002; Neumayer 2005). Moreover, Sikkink (2004, ch. 1) argues that the processes and rules adopted by individual states for ratifying treaties make a significant difference to whether treaties are ratified, meaning that where power is diffuse and the threshold for securing ratification is high, states are less likely to be able to ratify.

Third, some of our discussion of state compliance (or lack thereof) with international hate speech instruments in Chapter 4 overlaps with much broader debates about the relationship between international and domestic law. Most international law scholars see that international law is and should have supremacy over domestic law (see Glennon 1997; Goodman and Jinks 1997; Lowenfeld 1998; Stephens 1997). Even in the US context, for example, they can point to Art. VI of the US constitution: 'all Treaties made, or which shall be made, under the Authority of the United States, shall be the supreme Law of the Land'. Yet at the same time the US typically enters reservations, understandings and declarations (RUDs) to treaty ratification. For example, the US entered an understanding to the ICERD that the treaty was not self-executing: that it required domestic legal enactment in order to have domestic legal effect.[37] Such RUDs drain the binding force from treaties because they mean, for instance, that domestic courts cannot invoke the ICERD for purposes of directly enforcing its provisions against the wishes of the government. On the one hand, a potential benefit of RUDs is that they can make ratification possible, especially in contexts like the US where a super-majority is required for constitution change. More generally, some scholars believe that states can, and should, in some cases use RUDs in order to ensure that international law does not trump domestic law (Goldsmith 2000; Bradley and Goldsmith 2000; McGinnis and Somin 2006).

Fourth, some of what we discuss in Chapter 6 in relation to political arguments for and against hate speech laws, including campus speech codes, crosses over with wider debates on free speech in universities. Take the literature on the alleged snowflake generation, including cases of students demanding the removal of statutes and symbols of colonialism and requesting trigger warnings from lecturers on sensitive subject matter and content (see Fox 2016). Consider also the phenomenon of no platforming, especially in relation to speech deemed to be transphobic (see Heinze 2018; Srinivasan and Simpson 2018).

Fifth, some of what we say about the use of hate speech by political figures in Chapter 7 speaks to new insights on popularism. For example, many of the political figures we shall discuss have also been associated with what some scholars identify as popularism as 'a political style' (see Moffitt 2016), including the rhetorical style of making appeals to ordinary people. Closely allied to popularism as a political style is the presentational techniques of 'celebrity politicians', who make appearances at events that seem more like comedy clubs than political rallies and who use Twitter to directly communicate with voters, bypassing the traditional media (see Wood et al. 2016; Street 2018). At these events celebrity politicians combine together humour, argument, entertainment and propaganda. Importantly, many of the political figures we shall discuss in the book utilise hate propaganda in particular as

part of their popularism. This is designed to create or play on an existing sense of 'us' versus 'them' or 'the other', whilst identifying themselves and their intended audience (target voters) as the 'us'. Some of the political figures we shall discuss also attack judges, journalists, academics, public intellectuals, civil society organisations and, more generally, 'the liberal elite' as 'enemies of the people'. Whilst this might not be hate speech (depending on how this is defined), it nevertheless exemplifies, according to some scholars, an attempt by populist politicians to use middle- and working-class resentment against the professional classes for political gain (Wimberly 2018). According to other scholars, this is a worrying attempt to monopolise power by undermining other traditional sites of power (Moffitt 2016, ch. 8).

Sixth, some of the hate speech used by political figures we shall discuss in the book attacks economic migrants and asylum seekers using myths, false rumours and conspiracy theories. According to some scholars, political figures are able to get away with such falsehoods because in the relevant societies the dominant ways of ordinary people making sense of, appropriating and responding to their realities or lived experiences is not through an orientation of enlightenment rationality and intellectualism, but through an orientation of instinct, common sense, anti-intellectualism and irrationalism. The latter reaffirms and promotes more primitive tribal attachments to one's family, race and nation as lenses through which realities are perceived (Lanning 2012). In these societies appeals to both fear of the other and common sense can be discussion-ending trump cards (Saurette and Gunster 2011). This is part of a putative broader post-truth culture in public life. Populist politicians focus more on giving people 'the facts' that these politicians want the people to hear, and that the people want to hear, rather than the sorts of true facts which are supported by reports written by experts based on evidence (Block 2019). Indeed, voters may recognise (explicitly or implicitly) demagogic politicians as violating, even flagrantly, conventional norms of truth-telling yet nevertheless be willing to vote for them if they nonetheless perceive them as authentic champions of their interests, especially in the context of a perceived crisis of political illegitimacy (Hahl et al. 2018).

Seventh, some of what we discuss in the book, especially in Chapters 6 and 7, at times overlaps with academic literature on incivility and coarseness in public/political discourse. Pace Robert Post's (2009) assertion that hate speech laws are in the business of formalising civility norms, there are many important differences between the concepts of hate speech and uncivil speech (see Brown 2017a). Nevertheless, some of the academic debate on incivility in public life runs in parallel to debates about hate speech. In contrast to ubiquitous newspaper editorials by turns lamenting the coarsening of public/political discourse or ridiculing these very lamentations as thinned skinned political correctness, the academic literature on incivility offers characteristically more nuanced insights. For one thing, it is clear that incivility and coarseness in public/political discourse is not limited to countries with robust constitutional guarantees of free speech like America. Recent studies of public discourse in China, for example, reveals how its citizens often use rude or coarse language in creative ways – such as using the phrase 'ji de pi' (the fart of a chicken) as a substitute for gross domestic product (GDP) to express disagreement, disdain and anger towards government policy – to evade regimes of Internet censorship that remove overt political debate or criticism (see Wang 2018). For another thing, it is apparent that incivility is neither new nor noticeably worsening. In America, for example, angry and vituperative language was a notable fixture of political life in the 1990s (Esler 1997), as it was in the 1790s (Freeman 2001). More importantly, some scholars maintain that not merely has incivility always been a feature of American politics, and used to be much worse, but also that its presence should not be feared but viewed as a sign of a healthy, diverse ecosystem of debate. If there were no