This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Hate speech in the British press: a theoretical and practical assessment of the case for broader regulation

Hashim Ibrahim Hussein Mude

Presented for the degree of Doctor of Philosophy in 2022

Under the supervision of Dr Rachael Craufurd-Smith and Dr Paolo Cavaliere
Abstract

This thesis is concerned with the justifiability of regulating hate speech disseminated by the British press. It explains why the narrow scope of current regulation is at odds with prominent theories of press freedom, focusing on the following major accounts. Firstly, a libertarian theory which suggests the press ought to be subject to equivalent legal treatment to the public. Secondly, social responsibility theories which extend protection and special privileges to the press only when they fulfil certain democratic functions. It argues that these theories support broader regulation of hate speech than those which currently apply to the British press. Finally, it considers how such broader regulation can be implemented in a manner which preserves press freedom.

Chapter 1 explains the nature of hate speech in the British press identifying it as a form of covert hate speech that stereotypes and stigmatises racial, ethnic, and religious minorities. It explores why such content is not proscribed under the current legal regime (which only bans overt hate speech) and how it is caught by the broader restrictions applicable to the broadcast media. Finally, it explains that this gap is pertinent because hate speech in the media is mostly of the covert kind which escapes regulation.

Chapter 2 sets out the methodology used to explore the appropriate scope of hate speech restrictions on the press. It explains the coverage/protection distinction and the balancing framework that will be used to reconcile the press’s rights to publish with the need to prevent harm to others. It then explains circumstances under which such a balancing approach would support regulating covert hate speech under both the libertarian and social responsibility theories of press freedom.

Chapter 3 evaluates the rationales for the broader regulation of broadcasting, testing their coherency and whether they support the differential treatment of hate speech between broadcasting and the press.

In chapters 4 and 5, it is argued that regulation of covert hate speech by the press can be justified under the libertarian theory if it is shown to cause the same harms as overt hate speech. The chapters then set out the empirical evidence in the social science literature demonstrating a link between overt hate speech and various direct and
indirect harms. Finally, it considers whether there is evidence of similar quality linking covert hate speech by the press and these harms.

Chapter 6 examines theories of constituted harm built on J.L. Austin’s theory of speech acts. It explains the authority problem that prevents ordinary hate speakers from successfully ranking and subordinating their targets. It then argues that the press does not suffer from the same deficits in authority and considers the implications of this on the case for broader regulation.

Chapters 7 and 8 set out the conditions under which broader regulation of press hate speech would be supported by social responsibility theories. It is argued that such regulation would be justified if the press has more limited rights than the public to engage in hate speech and if such restrictions do not hinder their ability to perform their institutional functions. It is then contended that the strongest rationales for protecting hate speech in free speech theory are non-instrumentalist speaker-based arguments that either do not—or only marginally apply to—the press. Finally, it is argued that such regulation can be designed in ways that allow the press to perform their roles as public watchdogs, as a source of information and ideas on matters of public concern, and as a platform for people to exchange views.

Chapter 9 considers what regulation that is both effective against covert hate speech and which preserves press freedom would look like. This includes a consideration of whom it should apply, what form it ought to take, and how it would fit in the current regulatory framework.

This thesis, therefore, seeks to answer four main research questions:

1. What kind of hate speech does the press publish, and should it be restricted?
2. Does this hate speech cause harm, and if so, how?
3. What is the relationship between freedom of expression and press freedom, and how can these principles be reconciled with restrictions on covert hate speech?
4. What form should effective press-specific hate speech regulation take, who should it apply to, and where would it fit in the current regulatory framework?
Lay summary

This thesis considers whether negative stereotyping of racial, religious, and ethnic minorities by the British press should be regulated. It argues that this content ought to be restricted if it causes harm and sets out the evidence demonstrating that it does. It claims that owing to their communicative power and influence, the publication of such content by the press causes greater harm than hate speech published by other speakers.

It considers whether regulating this content is compatible with freedom of expression and press freedom. It does so by examining the various arguments in favour of a right to engage in hate speech, finding that these are weaker or do not apply to mass media institutions in the same way they do to other speakers. It is therefore argued that the press ought to have more limited rights to engage in hate speech and can be subject to broader restrictions. It also demonstrates that these restrictions do not affect the press’s ability to perform their democratic functions.

Finally, it considers how these restrictions could be implemented in practice by proposing regulation that is both effective at proscribing this content while leaving intact the press’s ability to shock, provoke, exaggerate, and offend. It also shows that such regulation is compatible with Article 10 of the European Convention on Human Rights.

The distinctiveness of this thesis lies in its consideration of whether more covert forms of hate speech ought to be regulated, a matter that is largely unaddressed in the existing literature. It differs from existing work in this area in its more in-depth analysis of empirical evidence from various disciplines, in its examination of underexplored theories of harm in the media law literature and in its discussion of how these restrictions would operate in practice.
Acknowledgements

I am deeply indebted to my supervisors, Dr Rachael Craufurd-Smith and Dr Paolo Cavaliere for their incredible support and patience throughout this journey. I am especially grateful to Rachael for her encouragement as I began this project, as well as her excellent Media Law course which inspired it and to Paolo for his invaluable career advice and guidance.

I would also like to express my deepest thanks to Dr Euan Macdonald for his thoughtful critique of my work, which enriched my thinking and writing on these topics. I am also thankful to the postgraduate research staff who masterfully navigated the Covid-19 pandemic and kept us going during unprecedented circumstances.

Special thanks to my parents, family and friends for their moral and emotional support. I am especially indebted to my partner, Elitsa. Much of this thesis was written up during the pandemic, and at times, her encouragement, belief in me and unconditional support were the only things keeping me going. For this, I will be forever grateful.

Finally, I would like to acknowledge the late Charles Edwin Baker. A remarkable thinker with plain and accessible writing whose work on free speech and media policy sparked my interest in these topics.
# Table of contents

Abstract 2

Lay summary 4

Acknowledgements 5

Chapter 1: Introduction 9

1. The purpose and scope of this inquiry 9

2. Introduction 12

3. Background of the problem 12

4. Definitions 14

   a. The press 15
   b. Hate speech: a whole spectrum of negative discourse 16

5. Dangerous negative stereotypes and covert hate speech 18

   a. Dangerous negative stereotypes 19
   b. The phenomena of covert hate speech 22

6. Hate speech laws and codes 27

   a. The purpose of press regulation 28
   b. The stirring up offences: limited to overt hate speech 29
   c. IPSO Editors’ Code of Practice 31
   d. IMPRESS Standards Code 36
   e. Hate speech regulation in broadcasting 39

7. Conclusion 44

Chapter 2: Methodology and Research Questions 45

1. Introduction 45

2. Theories of press freedom 46

   a. Libertarian theory and the equivalence model 46
   b. Social responsibility theories and the variable geometry model 48
   c. Bridging the divide 51

3. Methodology 53

   a. A note on comparative method 54
   b. Coverage 56
   c. Protection 58
   d. The conditions under which broader regulation is justified 66

4. Conclusion 67

Chapter 3: The Special Status of Broadcasting 68

1. Introduction 68

2. Is the special status of broadcasting justified? 68

   a. The scarcity of frequencies rationale 69
   b. The ‘best of both worlds’ rationale 73
   c. The intrusive medium rationale 76
   d. The special impact rationale 79

3. Conclusion 84

Chapter 4: The Direct Harms of Dangerous Negative Stereotypes 87

1. Introduction 87

2. Optimal balancing under the equivalence model 87
3. Hate speech and harm

4. Causing harm and empirical evidence: deterministic and probabilistic causation

5. Direct harms: Targeted and diffuse hate speech
   a. Psychological harm and severe-emotional distress
   b. Psychological harm in diffuse press hate speech: the absence of empirical evidence
   c. Misrecognition, self-esteem, and silencing
   d. The media’s unique capacity to silence: The phenomena of stereotype threat

6. Conclusion

Chapter 5: The Indirect Harms of Dangerous Negative Stereotypes

1. Introduction

2. The indirect harms of overt hate speech
   a. Violence and discrimination: contributing towards a climate of hatred
   b. Undermining security and dignity

3. Press hate speech and indirect harms
   a. DNS as a source of indirect harms
   b. Press hate speech: Priming and cultivating prejudice
   c. The effectiveness of moderate language

4. Conclusion

Chapter 6: Dangerous Negative Stereotypes and Constituting Harm

1. Introduction

2. Constituting harm and the authority problem

3. The limited authority of low types

4. The press as authoritative speakers

5. The consequences of constituting harm: enacting harmful social norms

6. Conclusion

Chapter 7: Social Responsibility Theories and a Free Speech Principle

1. Introduction

2. The SRT Challenge to hate speech regulation
   a. Optimal balancing under the variable geometry model

3. Non-instrumentalist speaker-based accounts for free speech
   a. Baker’s argument from self-disclosure: A right to be hateful
   b. Dworkin’s and Weinstein’s arguments from democracy: A citizen’s right to hateful political participation
   c. The inapplicability of speaker-based accounts to the press

4. A Non-instrumentalist audience-based account: Scanlon’s argument from audience autonomy

5. An Instrumentalist audience-based account: The argument from self-fulfilment

6. Conclusion

Chapter 8: Social Responsibility Theories and Press Freedom

1. Introduction

2. A principle of press freedom: coverage and protection
3. The watchdog role and hate speech .................................................. 184
4. Imparting hateful information and ideas ........................................... 190
   a. Republican and deliberative democratic models .......................... 190
   b. The liberal pluralist and complex democratic models .................. 195
5. Providing a platform for another’s hate speech ................................. 199
6. Conclusion .................................................................................... 201

Chapter 9: The Content and Structure of Press-Specific Hate Speech Regulation ................................................. 202

1. Introduction ..................................................................................... 202
2. The content of press-specific hate speech regulation .......................... 202
   a. A specific provision covering Dangerous Negative Stereotypes ....... 205
   b. The DNS clause in practice .......................................................... 209
3. The structure of press-specific hate speech regulation ........................ 214
   a. Who should it apply to? .............................................................. 214
   b. What form should it take? ............................................................ 218
   c. Discretionary benefits and burdens: towards a robust incentive structure .......................................................... 223
   d. The need for statutory underpinning ............................................. 227
   e. The enforcement powers .............................................................. 229
4. Compatibility of regulation with positive law ..................................... 233
   a. Legitimate aim ............................................................................. 234
   b. Suitability .................................................................................... 235
   c. Minimal impairment .................................................................... 235
   d. Proportionality strictu senso ......................................................... 238
5. Conclusion ..................................................................................... 244

Conclusion ......................................................................................... 246

1. What kind of hate speech does the press publish and should it be restricted? 246
2. What is the relationship between press freedom and free speech, and how can these principles be reconciled with restrictions on covert hate speech? .............. 248
3. What form should effective press-specific hate speech regulation take? __ 255
4. To whom should it apply? And where would it fit in the current regulatory framework? ............................................................................................................. 256

Bibliography ....................................................................................... 260

1. Books .............................................................................................. 260
2. Articles ............................................................................................ 271
3. Table of cases .................................................................................. 286
Chapter 1: Introduction

1. The purpose and scope of this inquiry

James Curran and Jean Seaton compare press regulation in Britain to the movie Groundhog Day. The real and perceived failures of the UK press often lead to calls for government intervention. Royal Commissions are established, reforms are proposed, and many are ignored. Another major failure occurs, and just like Phil Connors, we all wake up and relive the same day over and over again.

This thesis focuses on one of these points of failure, what Lord Leveson referred to as the ‘discriminatory, sensational or unbalanced reporting’ of racial, religious, and ethnic minorities. It is a philosophical inquiry into what role (if any) that press regulation ought to play in addressing this problem and an examination of how such regulation would apply in practice.

Under the status quo, hate speech published by the press against groups is subject to the same limits as that published by a lone hate speaker. This position reflects a certain philosophical understanding of press freedom, one of equivalence where the press’s right to publish is ‘neither more nor less than that of the general public’ and ought to be subject to the same limits.

The press has free reign to publish discriminatory content against groups, provided they do not do so using extreme language which is either intended or likely to incite hatred. This matters because such overt hate speech is a relic, an old racism of a bygone era. Rather, the hate speech which appears in some sections of the press is more covert, couched in more moderate language. It is a form of negative stereotyping and stigmatisation which encourages or justifies prejudice.

This thesis examines whether regulating such content is compatible with freedom of expression and press freedom. It does so by examining the theories underlying these

---

3 Attorney-General v Guardian Newspapers [1990] 1 A.C. 109, 183
4 The terms free speech and freedom of expression will be used interchangeably in this thesis unless the context suggests otherwise.
principles and determining what they inform us about where the line between protected and unprotected speech ought to be drawn and whether—and to what extent—this existing boundary (with respect to the UK press) can be extended to include some forms of covert hate speech.

The legitimacy of laws that restrict hate speech qua hate speech is a long-running and lively debate amongst political theorists and legal scholars. Various reasons have been presented in defence of these laws which have been countered by critiques on the other side of the discussion. Various UK-based scholars, including Eric Barendt, Ivan Hare, Eric Heinze, Robert Mark Simpson and Jeffrey Howard, have explored the arguments for and against such restrictions. Others, including Mary Matsuda, Charles Lawrence, James Weinstein, Jeremy Waldron, and Robert Post, to name a few, write from an American perspective. Many of them explore general arguments for either regulating or de-regulating hate speech. Others consider more specific types of hate speech, such as that directed against religious groups or sexual minorities. These debates are far from settled; they have been had and will continue to be had. This thesis will not be making a novel contribution to this literature but rather examines how these arguments apply to mass media institutions.

5 There is a distinction between content-based and content-neutral restrictions on hate speech, this thesis is concerned with the former Eric Heinze, Hate Speech and Democratic Citizenship (Oxford University Press 2016) ch 1.
7 Ivan Hare, ‘Crosses, Crescents and Sacred Cows: Criminalising Incitement to Religious Hatred’ [2006] Public Law 521.
8 Heinze, Hate Speech and Democratic Citizenship (n 5).
It is a contribution to a narrower body of literature which theorises press freedom and examines its relationship to freedom of expression. As Paul Wragg notes, this is something that ‘we have never gotten to grips with’. Eric Barendt, Helen Fenwick and Gavin Phillipson, and more recently, Paul Wragg and Damian Tambini, have all examined the nature of press freedom and the extent to which press claims for special immunities and privileges are justified.

My focus is different from theirs. I consider whether imposing broader hate speech restrictions on the press than those which currently apply to them is compatible with press freedom. Existing inquiries on hate speech in the media by the likes of Jacob Rowbottom and Jan Oster focus on overt hate speech. The justifiability of restricting covert hate speech and the form that such regulation ought to take are therefore novel contributions to the media law literature.

Finally, I explore the harms of hate speech by drawing on the empirical evidence from a variety of disciplines in the social sciences, including social and cognitive psychology, behavioural economics and communication studies. This builds on existing work by Mary Matsuda, Richard Delgado, Charles Lawrence and Jean Stefancic, as well as Alexander Brown’s analysis of hate speech laws which includes perhaps the most thorough examination of the empirical evidence on the harms caused by hate speech in the legal literature. I supplement this work both by focusing on the harms caused by covert hate speech which is under-explored in the legal literature, and by providing a more up-to-date account of the empirical evidence.

---

19 Barendt, *Freedom of Speech* (n 6).
23 Jacob Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’, *Extreme Speech and Democracy* (Oxford University Press 2009).
2. Introduction

The research questions and methodology that I will use to answer them are set out in chapter 2. Here I identify the nature of the problem, pointing out the kind of hate speech published by the press and why such content is not regulated by current press codes.

I start off in section 3 with a brief background of the problem of discriminatory coverage of minority groups by the British press. Section 4 defines how certain terms will be used in this thesis. Section 5 explains that the discriminatory content that is of concern is a form of negative stereotyping that encourages or justifies prejudices, what I term dangerous negative stereotypes. It explains the phenomena and sets out how and why the manner in which they are expressed has shifted from overt racism relying on tropes about biological inferiority to covert racism relying on highly group-specific cultural stereotypes. Section 6 explores how and why this content is not regulated under the current press codes but is caught by the broadcasting code.

3. Background of the problem

Some parts of the UK press have long been accused of engaging in discriminatory reporting about various racial, ethnic, and religious minorities, from the Jews and the Irish in the 19th and early 20th century, to gay people in the 50s and African-Caribbean and West Indians in the 70s and 80s. Scholars have noted that most of the rhetoric in the 21st century has been directed at Muslims, migrants, asylum-seekers, and traveller communities. Curran and Seaton recount several such examples, including an article describing Romanian migrants ‘boasting of plans to beg and steal’ and another suggesting Britain was engaging in ‘assisted suicide’ by taking in migrants and allowing the ‘spread of sharia law’. Human rights bodies have also taken note of this coverage. The UK is a signatory to the International Convention on the Elimination of all Forms of Racial Discrimination

27 Gordon Conway, Islamophobia: A Challenge for Us All (Runnymede Trust 1997); Elizabeth Poole, Reporting Islam: Media Representations of British Muslims (IB Tauris 2002); Chris Allen, Islamophobia (1st edn, Routledge 2010).
28 Curran and Seaton (n 26) ch 8.
('ICERD'), which requires states to proscribe hate speech.\textsuperscript{29} The ICERD committee monitors the implementation of the convention and produces periodic reports for each of its signatories. Similar work is done by the European Commission on Racism and Intolerance ('ECRI'), a body formed by the Council of Europe and tasked with monitoring racism and intolerance in the member states.

The ECRI's first report on the UK described 'widespread hostile sentiments' against the Muslim community evidenced by 'negative stereotypes in the media'.\textsuperscript{30} The second referred to 'racially inflammatory material' appearing in the print media.\textsuperscript{31} The third suggested that the tone of the media coverage of ethnic minorities was 'racist and inflammatory'.\textsuperscript{32} The fourth makes similar comments but goes further, suggesting a potential link between this coverage and violence meted out to these groups.\textsuperscript{33} Similar observations were made by the ICERD in its 2003 report on the UK which pointed to 'increased racial prejudice' against these groups in the media and the 'lack of effectiveness of the Press Complaints Commission in dealing with this issue'.\textsuperscript{34} Its 2011 report referred to 'virulent statements in the media that may adversely affect racial harmony and increase racial discrimination'.\textsuperscript{35}

In the same year, it was reported that journalists working at the \textit{News of the World} newspaper had engaged in phone hacking.\textsuperscript{36} The outrage from this scandal led to the formation of a public inquiry into press ethics led by Lord Leveson ('Leveson Inquiry'). Though the Leveson inquiry was ostensibly informed by the phone hacking scandal, its remit expanded to general considerations of press ethics. Jonathan Heawood points out that Leveson took particular interest in the issue of 'discriminatory

\begin{flushleft}
\textsuperscript{29} International Convention on the Elimination of All Forms of Racial Discrimination 1965, art 4.
\textsuperscript{30} ‘First Report on The United Kingdom’ (European Commission Against Racism and Intolerance 1999) 8.
\textsuperscript{34} ‘Concluding Observations of the Committee on the Elimination of Racial Discrimination (UK)’ (Committee on the Elimination of Racial Discrimination 2003) ICERD/C/63/CO/11 3.
\textsuperscript{35} ‘Concluding Observations of the Committee on the Elimination of Racial Discrimination (UK)’ (Committee on the Elimination of Racial Discrimination 2011) ICERD/C/GBR/CO/18-20 2.
\end{flushleft}
reporting’. The inquiry took evidence from a variety of minority groups and concluded that certain sections of the press were engaging in prejudicial reporting on matters concerning ethnic minority groups, with Leveson recommending that any new regulator ought to address this issue ‘as a matter of priority’.

The Leveson inquiry ended a decade ago, but scholars point out that these issues continue to plague the industry. This was evident during the ‘migrant crisis’ where comparative studies amongst European countries found the UK press to be uniquely hostile towards migrants, particularly those who were perceived to be Muslim. The ICERD committee’s 2016 report expressed concerns that the ‘negative stereotyping and stigmatization’ of these groups in the media remained unaddressed. The ECRI’s fifth report echoed similar sentiments describing the print media coverage as ‘hate speech’. Notably, all these reports indicate that this problem is largely confined to the print media. It will be noted that this is because of the more robust regulation of such content that is applied in the broadcast sector.

4. Definitions

Before examining the legal framework which applies to such content, I will first set out how the terms press, hate speech, hate speech laws, and hate speech codes will be used in this thesis.

---

38 Leveson (n 2) 673.
a. The press

Media law scholars point to two ways of defining the press. The first is an institutional definition which covers traditional mass media institutions such as the printed press.\textsuperscript{42} The second is functional, which includes anyone engaging in journalistic activities.\textsuperscript{43} There are good reasons to prefer the functional definition if one is considering whether certain speakers ought to benefit from the coverage and protection of a principle of press freedom.\textsuperscript{44} I will argue that in such cases the only question should be whether the activity they are engaging in (either in newsgathering or reporting) fulfils any of the functions that justify protecting press freedom in the first place, what I term the press’s democratic functions.\textsuperscript{45} Such protection ought to apply not to media institutions as \textit{institutions} but rather to \textit{any speaker} performing these functions.\textsuperscript{46}

However, this thesis is concerned with press regulation and this requires specifying the institutions they will apply to. As Rowbottom points out, publishers would need to know in advance that they are subject to a certain system and standards.\textsuperscript{47} It is also concerned with questions of harm, and this harm is largely a consequence of communicative power.\textsuperscript{48} The focus will therefore be on institutions such as the national and regional press that possess it. Much has been written about the plummeting sales of physical papers, the shrinking of advertising budgets and the hollowing out of newsrooms.\textsuperscript{49} But even in such a climate, the institutional press continues to wield significant communicative power. In the UK, 40\% of UK adults receive their news from the print or online versions (websites and apps) of these newspapers.\textsuperscript{50} Many of the other popular news sources, such as social media platforms including \textit{Meta}, increasingly rely on and amplify ‘trusted news’ sources and original reporting from

\textsuperscript{42} Jacob Rowbottom, \textit{Media Law} (Hart Publishing 2018) 38; Peter Coe, ‘Redefining “Media” Using a “Media-as-a-Constitutional-Component” Concept: An Evaluation of the Need for the European Court of Human Rights to Alter Its Understanding of “Media” within a New Media Landscape’ (2017) 37 Legal Studies 25, 27.
\textsuperscript{43} Rowbottom, \textit{Media Law} (n 42); Coe (n 42) 37–41.
\textsuperscript{44} Rowbottom, \textit{Media Law} (n 42) 31.
\textsuperscript{45} See chapters 2, 8 and 9.
\textsuperscript{46} Coe (n 42) 41–49.
\textsuperscript{47} Rowbottom, \textit{Media Law} (n 42) 31.
\textsuperscript{48} See chapters 4, 5 and 6 for a discussion of the various direct, indirect, and constituted harms.
institutional actors.\textsuperscript{51} They are the agenda-setters. The editorial choices of the institutional press, what they decide to report (and not report) and how they do it frame the terms of the political debate.\textsuperscript{52}

But power is not the preserve of the institutional press. One of the critiques of the Leveson inquiry was its failure to account for prominent bloggers and other forms of ‘new media’.\textsuperscript{53} I will therefore consider the normative justification of applying such regulation to forms of new media. I will also set out the criteria that ought to be used to determine which publications to include in the regulatory framework. Finally, this thesis is concerned with identifying the gaps in the current press codes which inform the failure to deal with the issue of discriminatory reporting. This means focusing on the institutional press as these are the bodies currently subject to such regulation.

b. Hate speech: a whole spectrum of negative discourse

Hate speech is a notoriously vague term that does not admit of easy definition as it is widely deployed to refer to what Tarlach McGonagle describes as a ‘whole spectrum of negative discourse stretching from hate and incitement to hatred; to abuse, vilification, insults and offensive words and epithets; and arguably also to extreme examples of prejudice and bias’.\textsuperscript{54} A detailed examination of this entire spectrum is a mammoth endeavour that is not feasible within a single thesis. As Alexander Brown notes, most legal scholars have examined ‘one genus, one species or even one instance’ of hate speech.\textsuperscript{55} Further, this thesis is not a general examination of the phenomena of hate speech. It is only concerned with the genera relevant to the press. The first genus is what Brown classifies as ‘negative stereotyping and stigmatisation’.\textsuperscript{56} It is widely accepted that this genus is the most relevant to the press. This has been

\textsuperscript{51} James Meese and Edward Hurcombe, ‘Facebook, News Media and Platform Dependency: The Institutional Impacts of News Distribution on Social Platforms’ (2021) 23 New Media & Society 2367;\
\textsuperscript{52} Wragg, \textit{A Free and Regulated Press} (n 18) 17; Tambini (n 22) ch 1; Curran and Seaton (n 26) ch 26.\
\textsuperscript{53} Wragg, \textit{A Free and Regulated Press} (n 18) 283.\
\textsuperscript{54} Tarlach McGonagle, ‘Minorities and Online “Hate Speech”: A Parsing of Selected Complexities’ (2012) 9 European Yearbook of Minority Issues Online 419, 420.\
\textsuperscript{55} Alexander Brown, \textit{Hate Speech Law: A Philosophical Examination} (1st edn, Routledge 2015) 19.\
\textsuperscript{56} ibid 21–22.
noted in scholarship,\textsuperscript{57} by the Leveson inquiry,\textsuperscript{58} and in the ECRI and ICERD reports referred to above. I explore it in greater detail in section 5 below.

The second genus is incitement to hatred which refers to expressions which stir up hatred against groups based on certain ascriptive characteristics.\textsuperscript{59} This genus lies at the heart of much of the hate speech debates in legal scholarship where the term hate speech is often used as a synonym for incitement to hatred.\textsuperscript{60} It is also expressly referred to in various international human rights conventions, including the ICERD and the International Covenant on Civil and Political Rights (‘ICCPR’).\textsuperscript{61} Most importantly, it is currently the only genus of hate speech targeting groups regulated in the press. This is also set out in greater detail in section 6 below.

The wide spectrum of hate speech is regulated by a patchwork of criminal offences in the UK (‘hate speech laws’). These include ‘aggravated’ offences where existing non-speech-related crimes have higher maximum penalties if they contain an element of hostility based on race or religion,\textsuperscript{62} ‘racialist chanting’ at football matches\textsuperscript{63}, the offences of harassment, intentional harassment and causing fear or provocation of violence under the Public Order Act\textsuperscript{64} as well as communications offences relating to abusive and grossly offensive content.\textsuperscript{65} The hate speech laws I will focus on are the offences of stirring up hatred (‘stirring up offences’) with respect to race, religion and sexual orientation,\textsuperscript{66} as this is the cluster relevant to incitement to hatred.


\textsuperscript{58} Leveson (n 2) 665–673.

\textsuperscript{59} Brown, \textit{Hate Speech Law} (n 55) 26.

\textsuperscript{60} See generally Ivan Hare (ed), \textit{Extreme Speech and Democracy} (Oxford University Press, USA 2011); Michael Herz and Peter Molnar (eds), \textit{The Content and Context of Hate Speech: Rethinking Regulation and Responses} (Cambridge University Press 2012); Waldron, \textit{The Harm in Hate Speech} (n 14); Heinze, \textit{Hate Speech and Democratic Citizenship} (n 5).

\textsuperscript{61} International Covenant on Civil and Political Rights 1966, art 20(2).

\textsuperscript{62} Crime and Disorder Act 1998, s 28-32.

\textsuperscript{63} Football (Offences) Act 1991, s 3.

\textsuperscript{64} Public Order Act 1986, s 4, 4A, and 5.

\textsuperscript{65} Communications Act 2003, s 127; Malicious Communications Act 1988, s 1.

\textsuperscript{66} Public Order Act 1986, prt 3 and 3A.
Hate speech laws often regulate a narrow subset of speech within a genus of hate speech because the law attempts to strike a balance between prohibiting hate speech and respecting fundamental rights. \(^{67}\) For example, incitement to racial hatred (hate speech) is only prohibited in the UK where it is committed using words, behaviour or written material which are threatening, abusive or insulting (hate speech law).\(^{68}\) Finally, there is also media-specific regulation containing provisions relevant to this subset of hate speech (‘hate speech codes’). This includes the Ofcom Broadcasting Code (‘Broadcasting Code’) and the codes of conduct applied by press regulators.

I will now examine the nature of stereotypes and identify the specific kind of stereotyping that is of concern to the scholars and various institutions set out above.

**5. Dangerous negative stereotypes and covert hate speech**

Social psychologists claim that ‘inherent to human cognition’ is the emotional need to be part of a group and to categorise those within it (ingroups) and outside it (outgroups), creating *us* and *them*.\(^{69}\) People are placed into outgroups based on perceived commonalities and then stereotyped by indiscriminately ascribing to them ‘common psychological characteristics’.\(^{70}\)

The media do not create stereotypes but are a principal form by which they are maintained and circulated.\(^{71}\) Stereotypes are group-specific; Catholics are superstitious; Indians like spicy food etc. They can assign positive traits; group *x* is hardworking or negative ones; group *y* is lazy. Some argue that even expressing positive stereotypes is problematic because we are then assumed to also believe the negative ones.\(^{72}\) Others argue that all stereotypes are inherently negative because they fail to treat people as individuals.\(^{73}\) The term ‘negative stereotypes’ can therefore

---


\(^{68}\) Public Order Act 1986, s 18-19.


\(^{72}\) Nelson (n 57) 4.

\(^{73}\) Ibid.
refer to a broad range of content. My focus is on a specific type of negative stereotyping.  

a. Dangerous negative stereotypes

The principal concern expressed by scholars, human rights bodies and during the Leveson inquiry is the potential for discriminatory press coverage to cause harm; by intimidating groups, creating an atmosphere of hostility and rejection, inciting racial hatred, fuelling discrimination, and damaging community cohesion. It is in light of this that Lord Leveson made the following recommendation:

*I recommend that consideration should also be given to Code amendments which, while fully protecting freedom of speech and the freedom of the press, would equip that body with the power to intervene in cases of allegedly discriminatory reporting, and in so doing reflect the spirit of equalities legislation.*

Jonathan Heawood and Brigit Morris have interpreted this recommendation (‘Recommendation 38’) as a reference to the harassment provisions under section 26 of the Equalities Act, 2010, arguing that Leveson wanted a new regulator to consider proscribing press content that creates an ‘intimidating, hostile, degrading, humiliating or offensive environment’ based on a protected characteristic.

However, the harassment provisions are but one example of discriminatory conduct. The act’s purpose—or spirit—is broader and is set out in its explanatory notes, which is to protect people from ‘discrimination, harassment or victimisation’. It is clear from the report that Leveson’s principal concern was about press content ‘fostering prejudice’ and the potential for this to cause discrimination. The concern then is about negative stereotyping which encourages or justifies prejudice, what I term Dangerous Negative Stereotypes (‘DNS’). I turn to them now.

---

74 Brown, *Hate Speech Law* (n 55) 22.
75 *Concluding Observations of the Committee on the Elimination of Racial Discrimination (UK)* (n 35) 2; Leveson (n 2) 473–490; *ECRI Report on the United Kingdom (Fifth Monitoring Cycle)* (Council of Europe 2016) CRI(2016)38 21.
76 Leveson (n 2) 37.
77 Heawood and Morris (n 37) 31.
78 Leveson (n 2) 478, 490, 665-666,670-673.
I. Essentialism

Erik Bleich writing in the context of the Danish Cartoon Affair,79 stated that a ‘line is crossed when criticism evolves into essentializing, stereotyping, and branding the entire group as dangerous or inferior with the likelihood of stirring hatred’.80 Here he identifies the two components of DNS; essentialism and assigning an undesirable trait.

Gordon Allport, in his landmark text The Nature of Prejudice distinguishes between two ways of categorising and stereotyping. The first are differentiated categories where beliefs are more flexible, allowing for variation between members of the outgroup; ‘Catholics are generally superstitious, but many are not’.81 They are more tentatively held and are open to change when encountering new information.82

The second are monopolistic categories. They are more rigid, and the outgroup is undifferentiated. These are often defended when met with conflicting evidence, and it is ‘difficult to disabuse people of them’.83 Essentialism is a type of monopolistic categorisation. Essences define the very nature of a group, most of them have it, and it is unchanging; ‘Jews are just born greedy’.84 These pathologies are then used to justify the outgroup’s subordinate position in the social hierarchy.85 What makes essentialism dangerous therefore is the unchanging nature of these essences, their attribution to most (or all) of the outgroup and the difficulty in disabusing people of them.

II. Undesirable traits

The second feature of DNS is the nature of the traits attributed to groups. The concern is about those ‘deficiencies’ that offer a ‘strong reason to exclude them from social and political life and/or give their interests less consideration’.86 Recall that stereotypes are group-specific, and therefore these undesirable traits will differ between groups, e.g.

81 Allport (n 69) 173.
83 Jost and Hamilton (n 70) 219.
84 Allport (n 69) 173; Parekh, Rethinking Multiculturalism (n 57) 217.
85 Jost and Hamilton (n 70) 219.
Black people as unintelligent and prone to criminality, Jews as controlling society and culture, Roma as illiterate and lazy, Muslims as dangerous and uniquely prone to violence.

These undesirable traits have been referenced by several scholars in the hate speech literature. Brown writes about virulent forms of stereotyping that misattribute negative traits such as a lack of intelligence or those which stigmatize a particular characteristic. Mary Matsuda notes the various ills caused by ‘racist hate propaganda’, which is a euphemism for a particular set of pernicious racial stereotypes. Parekh writes of ‘malicious stereotypes’ used to justify the mistreatment of Jews and Black people in Europe.

The kinds of undesirable traits and the consequences that flow from them broadly map into the following categories. First are those traits which encourage hostility, such as claims that certain groups are dangerous or violent, e.g. Muslims are jihadists. Second are those which justify discrimination by characterising certain groups as inferior, unintelligent, or backward. Third are those which encourage contempt by painting certain groups as degenerate or deviant.

Such stereotyping is strongly linked to incitement to hatred. This is evident from the examples of incitement used by hate speech scholars. I explore these later, but Waldron, Tsesis and Parekh all refer to these undesirable traits in their examples of speech constituting incitement. This is because it is easier to convince people to hate

---

90 Allen (n 27).
91 Brown, Hate Speech Law (n 55) 133.
92 Matsuda (n 11) ch 2.
93 Parekh, Rethinking Multiculturalism (n 57) 316.
95 Matsuda (n 11) 25.
96 Brown, Hate Speech Law (n 55) 133.
97 See Chapter 5
others when the message relies on malicious stereotypes that are already associated with certain groups, as this takes advantage of existing cognitive biases.

b. The phenomena of covert hate speech

The nature of these undesirable traits has shifted over time. I will examine this briefly by setting out how and why overt racism and biological markers of inferiority were replaced by covert racism and cultural essentialism. There are several specialised texts that examine this phenomenon in Britain, including the work of Arun Kundnani, Reena Bhavnani, Heidi Mirza, and Veena Meetoo. Paul Hartmann and Charles Husband focus specifically on the mass media.

I. The Old Racism

Hartmann and Husband argue that racist discourse in the Elizabethan era was based on theological conceptions of other races. God created man to be ‘angel-like’ and ‘white’ whereas Black men were ‘barbaric, libidous and un-Christian’. Kundnani refers to this as a period of ‘unorganised racial beliefs’ with no clear delineation between different races with everyone who was not a White Briton lumped into the category of ‘moors’, which covered everything from Arab Muslims to Black Africans. It was only in the 17th and 18th-century era of the slave trade (and the need to provide intellectual justification for chattel slavery) that these racial categories became more rigidly defined. These ‘racial differences’ were then used to ‘define fitness for enslavement’.

Increased secularisation during the enlightenment and the ascendency of science marked a shift in the intellectual underpinnings of racist discourse. Theoretical explanations for racial hierarchies gave way to pseudo-scientific ones. The emergent disciplines of anthropology and biology were used to establish the supposed

---

101 Hartmann and Husband (n 100) 23.
102 ibid 22–23.
103 Kundnani (n 98) 11.
104 ibid 18; Romm (n 100) 44; Bhavnani, Mirza and Meetoo (n 99) 8.
105 Hartmann and Husband (n 100) 23.
‘physical’ grounds of racial differences which were linked to social and psychological differences between different racial groups. For example, Black Africans were stereotyped as predisposed to being ‘lazy and lacking ambition’. These DNS were not limited to this group. While hostility against Indians was not as vicious, they were still regarded as an inferior race which was conquered, and the Irish were considered as ‘literally dark’.

In this era, race was not only biological but innate and fixed. This changed with the emergence of Social Darwinism and the birth of eugenics which argued that race itself was subject to change and that the ‘civilised races’ could be ‘contaminated’ by mixing with others, particularly ‘lesser breeds’ which were unfit for survival and doomed to extinction. Lesser breeds were to be kept separate to avoid contaminating the populace. These ideas would provide the intellectual support for the elimination of entire populations and inspired the German Nazi regime’s program of ‘race purification’.

At this time, the British public viewed racism as natural and uncontroversial or, as Hartmann puts it, ‘it not only came to seem normal that blacks should be subordinate to whites, but the idea that things might be otherwise seemed unnatural, ludicrous or wrong’. These ideas were self-reinforced by a legal order that either explicitly or implicitly discriminated against ‘other’ races. As an example, the law in the 18th and early 19th centuries treated African slaves no differently from chattel and excluded Jews from various professions and public office on explicitly antisemitic grounds.

106 ibid.
107 ibid 26; Kundnani (n 98) 19.
108 Kundnani (n 98); Bhavnani, Mirza and Meetoo (n 99) 9.
110 Hartmann and Husband (n 100) 29.
111 While atheists and Catholics were equally persecuted on religious grounds Ursula Henriques notes Jews were excluded due to their ‘ancestors having crucified Jesus’, their alleged greater loyalty to their ‘own race’ and that they were of ‘bad character and morally inferior’ all grounds drawn from various antisemitic tropes Ursula Henriques, ‘The Jewish Emancipation Controversy in Nineteenth-Century Britain’ [1968] Past & Present 126, 126, 134.
Overt hate speech was entirely normal in this period. The only difference was that well-educated ‘elites’ made ‘fine distinctions’ between races, whereas the ‘ordinary Briton’ tended to think and speak of the ‘other’ as an undifferentiated mass.\textsuperscript{112}

The two key takeaways from this era are that racism was expressed overtly (‘overt hate speech’) using language we now consider to be extreme, and DNS were couched in biological terms.

\textit{II. New racism and cultural essentialism}

The post-war period saw a liberalisation of social attitudes towards minority groups which made it anathema to engage in overt racism.\textsuperscript{113} In 1965 parliament passed the Race Relations Act with the twin goals of ‘prohibiting discrimination on racial grounds in places of public resort’ and ‘penalising incitement to hatred’ (the precursor to the current stirring up offences).\textsuperscript{114} Gavin Schaffer points out that this law was targeted to deal with fringe extremists by making an ‘outright brand of racism and antisemitism...legally beyond the pale’.\textsuperscript{115} Commentators such as Anthony Dickey as well as Anthony Lester and Geoffrey Bindman argue that these laws had the effect not of banning racist speech but rather moderating the language and making it ‘more respectable’.\textsuperscript{116}

Hate speech laws are often critiqued on this basis by those who oppose content-based restrictions on speech (‘oppositionists’).\textsuperscript{117} Robert Post points out that they are ineffective because they cover the ‘rantings of a street-corner bigot’ but allow the same message to be expressed in more sophisticated language (‘covert hate speech’).\textsuperscript{118} Similarly, Dworkin notes that these laws do not cover the moderate expressions that

\textsuperscript{112} Hartmann and Husband (n 100) 27–28.
\textsuperscript{115} ibid 262.
\textsuperscript{117} This is a term borrowed from Heinze, oppositionists are a broad camp encompassing those who oppose viewpoint-based restrictions on speech while prohibitionists are in favour of them Heinze, \textit{Hate Speech and Democratic Citizenship} (n 5).
\textsuperscript{118} Post, ‘Hate Speech’ (n 15) 135; Ronald Dworkin, ‘Reply to Jeremy Waldron’ in Michael Herz and Peter Molnar (eds), \textit{The Content and Context of Hate Speech: Rethinking Regulation and Responses} (Cambridge University Press 2012) 343.
are ‘actually used to inspire hatred’. Edwin Baker and Eric Heinze make the same observations.

For the same reasons, the kinds of DNS used have become more moderate. Shortly after the full scale of the horrors of the Holocaust became clear, there was a concerted effort by the academic community to debunk the eugenicist theories of biological inferiority that were core to the Nazi dehumanisation of Jews. The 1950s and 60s saw a variety of UNESCO statements on this topic, including the seminal paper the Race Question, which argued that race is a social rather than a biological construct, and therefore claims of biological inferiority were both morally and scientifically unsound, a position which reflects the consensus in the literature today.

This old racism ended, and it was replaced by a ‘new racism’. This is a term coined by Martin Barker to describe the language and rhetorical techniques used in racist discourse in the post-war period. Various scholars have used different terms to describe this and related phenomena, including ‘cultural racism’, ‘symbolic racism’ and ‘modern racism’. At the core of theories of new racism is the recognition that while the theological and biological conceptions were dominant in old racist discourse, minority groups are now pathologized along cultural lines using more moderate language. Kjartan Páll Sveinsson suggests that this is because culture is seen ‘as a non-biological concept, it is supposedly nonracialized and thereby non-racist’. Whereas one would be perceived as an extremist for attacking minorities because of their alleged biological inferiority, they could escape such a label by

---

119 Dworkin, ‘Reply to Jeremy Waldron’ (n 118) 343.
121 Heinze, Hate Speech and Democratic Citizenship (n 5) 29.
122 Ashley Montagu, Man’s Most Dangerous Myth: The Fallacy of Race (Rowman & Littlefield 1997); Bhavnani, Mirza and Meetoo (n 99) 10.
attacking them based on differences in culture. The linguist Teun Van Dijk points out how cultural racism manifests by noting:

_They have a different culture, although in many respects there are deficiencies, such as single parent families, drug abuse, lacking achievement values, and dependence on welfare and affirmative action—'pathologies' that need to be corrected of course._ 129

Kundnani makes similar observations noting that ‘in place of overt statements of racial superiority came the idea of innate cultural differences between ethnic groups’ and while before immigration had to be stopped to ‘preserve the racial health’ of the population, now it had to be halted to preserve national identity from being corrupted by inferior cultures. 130 Of note is that while these new pathologies are presented as cultural deficiencies, they were still regarded as innate and unchanging. 131 The old biological DNS were simply repackaged as cultural ones. As Kundnani notes, Asian culture made them ‘obstinately backwards’ and gave them a ‘hysterical political temperament’. 132 Britons from an African-Caribbean background were suffering from a ‘flawed family structure and values’, and ‘lacked educational ambition’. 133 New racism also manifested in ‘elite discourse’, a term used to refer to an even more ‘articulate, moderate and superficially sophisticated racism’ deployed by politicians, academics, and the media. 134 Kundnani notes that a renowned academic argued that the Brixton riots in London were due to the African-Caribbean’s ‘innate cultural propensity to disorder, even primitive fascination with fire’. 135 Similarly, Hartmann and Husband, Paul Gordon and David Rosenberg, Van Dijk, and more recently, Gavin Titley identify these rhetorical strategies and styles of new racism in the British press and their use of cultural DNS. 136

---

130 Kundnani (n 98) 43.
131 Sveinsson (n 2) 3.
132 Kundnani (n 98) 44.
133 ibid 44.
135 Kundnani (n 98) 44.
136 Hartmann and Husband (n 100); Paul Gordon and David Rosenberg, _Daily Racism: The Press and Black People in Britain_ (Runnymede Trust 1989) chs 2–5; Dijk (n 134) ch 7; Gavan Titley, _Racism and Media_ (SAGE 2019) ch 2.
Unlike DNS based on biological markers, DNS based on culture are extremely group-specific, with Asians, Black people, Jews, eastern Europeans, Muslims etc., all subjected to and stereotyped with distinct cultural traits. These DNS can also manifest in ways that are societal or even region-specific. For example, Chris Allen notes that Islamophobia manifests in subtly different ways and utilises different tropes depending on the society.137

The key takeaway from this is that due to legal prohibitions and changing social attitudes, racist discourse shifted from banned and crude overt hate speech to lawful covert hate speech couched in moderate, coded and more sophisticated language. For the same reasons, ethnic minorities are now pathologized based on group-specific cultural DNS rather than those based on biological differences.

This is the form that content encouraging or justifying prejudice takes in contemporary Britain. Therefore, this is the type of content which would have to be regulated if Leveson’s vision of proscribing discriminatory reporting is to be realised. In this next section, I examine the relevant press codes setting out how and why they do not cover such content. Thereafter, I will demonstrate that this content is covered and proscribed by the Broadcasting Code.

6. Hate speech laws and codes

The UK press have not been subject to any press-specific statutory regulation for hundreds of years (since the expiry of the Licensing of Press Act in 1695) but instead self-regulate under the umbrella of various voluntary press organisations.138 Press misconduct has often led to the formation of public inquiries to investigate standards, often leading to the press disbanding their existing regulator and forming a new one. The Press Council was replaced by the Press Complaints Commission ('PCC') after the 1989 Calcutt inquiry found it was ineffective, lacked independence and had insufficient enforcement powers.139 This was then followed by the Leveson inquiry mentioned earlier, which found that similar to the Press Council before it, the PCC was ‘neither an effective not independent regulator’, with Leveson recommending that it be

---

137 Allen (n 27) 105.
139 Tom O’Malley and Clive Soley, *Regulating the Press* (Pluto Press 2000) 88
replaced with a system of ‘voluntary independent self-regulation’ underpinned by statute.\textsuperscript{140}

However, largely due to press lobbying, Leveson’s recommendations for statutory underpinning and measures to incentivise membership were watered down.\textsuperscript{141} Instead of a statute, the government established a Royal Charter on Self-regulation of the Press (‘Royal Charter’) which created the Press Recognition Panel (‘PRP’), a body tasked with recognising and approving press regulators.\textsuperscript{142} Regulators would now have to apply to the PRP for recognition and had to satisfy the panel that they met the criteria specified under the Royal Charter, which were drawn from some of Leveson’s recommendations.\textsuperscript{143} These include effectiveness, transparency, independence, sufficient enforcement powers etc.\textsuperscript{144} Recommendation 38, which, as noted above, required regulators to have the power to deal with discriminatory content, was not included as part of the mandatory recognition criteria.

\textbf{a. The purpose of press regulation}

The press is subject to existing civil and criminal law. Therefore, a regulatory body can serve either of the following purposes.\textsuperscript{145} One purpose is to set and enforce content standards on the press that go beyond those under general civil and criminal law. Both major press codes explored below have such provisions, including requirements to avoid inaccuracies,\textsuperscript{146} to not intrude into grief,\textsuperscript{147} and to be careful about reporting details of suicide and self-harm.\textsuperscript{148}

The second purpose is what Leveson terms ‘regulatory enforcement’, which means setting content standards that do not go beyond existing civil and criminal law. Here the regulatory framework only serves to provide complainants with a cheap and

\begin{itemize}
\item \textsuperscript{140} Barendt and others (n 138) 44; Paul Wragg, A Free and Regulated Press: Defending Coercive Independent Press Regulation (Hart Publishing 2020) 264–265.
\item \textsuperscript{141} See discussion in Chapter 9.
\item \textsuperscript{142} Royal Charter on Self-regulation of the Press 2013, cl 4.
\item \textsuperscript{143} Ibid sch 3.
\item \textsuperscript{144} Ibid; Paul Wragg, ‘Leveson’s Vision for Press Reform: One Year On’ (2014) 19 Communications Law 6, 8.
\item \textsuperscript{145} Robert Baldwin, Martin Cave and Martin Lodge, Understanding Regulation: Theory, Strategy, and Practice (Oxford University Press 2011) ch 1.
\item \textsuperscript{146} IPSO Editors’ Code 2021, cl 1; IMPRESS Standards Code 2017, cl 1.
\item \textsuperscript{147} IPSO Editors’ Code 2021, cl 4; IMPRESS Standards Code 2017, cl 7.2.
\item \textsuperscript{148} IPSO Editors’ Code 2021, cl 5; IMPRESS Standards Code 2017, cl 9.
\end{itemize}
accessible avenue to seek redress against the press for matters which are already proscribed under the general law without having to resort to expensive litigation, e.g. an arbitration scheme to pursue breach of privacy or defamation claims.\textsuperscript{149} This second approach is what the independent Monitor for the Press (‘IMPRESS’) has taken regarding hate speech against groups with their code mirroring the stirring up offences. IMPRESS argues that:

\textit{while hate speech offences exist in criminal law, we consider Clause 4.3 provides access to justice to members of the public, as well as acting to increase awareness of the issue in the print media.}\textsuperscript{150}

The Independent Press Standards Organisation (‘IPSO’) takes neither approach with its code failing to address the matter entirely. It does not set standards that go beyond the general law, and neither does it mirror it, leaving the matter entirely to the criminal justice system. Therefore, either through mirroring the standards under the general law (IMPRESS) or leaving the matter to it (IPSO), the press can publish hate speech against groups as long as it does not breach the stirring up offences. Therefore, before examining these press codes, I will first set out the type of content covered by these offences.

b. The stirring up offences: limited to overt hate speech

There are two categories of stirring up offences in England and Wales. The first is the broader offence of stirring up racial hatred, which proscribes the use of threatening, abusive or insulting material intended or likely to stir up hatred (‘broader stirring up offence’).\textsuperscript{151} Racial hatred includes nationality as well as certain groups that are considered to be ‘racial’, such as Jews and Sikhs.\textsuperscript{152}

Multi-ethnic groups such as Muslims are treated only as religious groups and are covered by the narrower second category of offences that proscribe stirring up hatred

\begin{itemize}
\item \textsuperscript{149} Leveson (n 2) 1784–1785.
\item \textsuperscript{151} Public Order Act 1986, s 18-19.
\item \textsuperscript{152} Kay Goodall, ‘Incitement to Religious Hatred: All Talk and No Substance?’ (2007) 70 The Modern Law Review 89, 90.
\end{itemize}
against people based on their religion or sexual orientation (‘narrower stirring up offences’).\(^\text{153}\) These offences are narrower because they are limited only to the use of threatening material and specific intent rather than a likelihood that hatred will be stirred up. The narrower stirring up offences contain a clause protecting freedom of expression with respect to criticism of religious beliefs, which provides:

> Nothing in this Part shall be read or given effect in a way which prohibits or restricts discussion, criticism or expressions of antipathy, dislike, ridicule, insult or abuse of particular religions or the beliefs or practices of their adherents, or of any other belief system or the beliefs or practices of its adherents, or proselytising or urging adherents of a different religion or belief system to cease practising their religion or belief system.\(^\text{154}\)

It also includes a similar clause with respect to criticism of sexual orientation:

> the discussion or criticism of sexual conduct or practices or the urging of persons to refrain from or modify such conduct or practices shall not be taken of itself to be threatening or intended to stir up hatred…for the avoidance of doubt, any discussion or criticism of marriage which concerns the sex of the parties to marriage shall not be taken of itself to be threatening or intended to stir up hatred.\(^\text{155}\)

There is also a distinction between the stirring up offences in England and Wales and the one in Scotland. In Scotland, there is a single offence which replicates the broader stirring up offences in England and Wales but which captures all the protected characteristics set out above and adds age as well as transgender identity.\(^\text{156}\) A core ingredient of all these offences is that there must be a stirring up element, and the hate speech must be expressed in an abusive, insulting, or threatening manner (‘qualifiers’). The court has confirmed that the qualifiers must be given their ‘natural and ordinary meaning’.\(^\text{157}\) Abusive and insulting material is generally regarded as that which is derogatory in nature and includes the use of profane or extreme language,

\(^{153}\) Public Order Act 1986, s 29C.
\(^{154}\) ibid s 29J.
\(^{155}\) ibid s 29JA.
\(^{156}\) This law is not in force at the time of writing Hate Crime and Public Order (Scotland) Act 2021, s. 4
such as racial epithets. Moderate expression of hatred without resorting to abuse or insults would not breach the law.

Threatening on the other hand, is even narrower. For example, the holding up of anti-gay leaflets titled ‘G.A.Y. God Abhors You’ and ‘Turn or Burn’ does not meet this requirement. Rather the threshold into ‘threatening’ is crossed when the material asserts the victims will or should be harmed in this life—not the afterlife—such as the suggestion that the ‘death sentence was the only way that the crime (of homosexuality) can be erased from society’. Similarly, expressions such as ‘Bomb a mosque day’ and ‘Kill a coon’ would constitute threatening words. Commentators have long noted that these offences are so restrictive as to lead to a failure to prosecute anything short of the most overt forms of hate speech. So much so that the recent review of hate speech laws by the Law Commission of England and Wales (‘Law Commission’) has recommended removing the qualifier requirement in some cases. This explains why they are only a handful of prosecutions for either of these offences every year. I will now examine the codes applied by the two regulators and set out how these narrow restrictions affect the ability to restrict DNS by the press.

c. IPSO Editors’ Code of Practice

Most national newspapers are regulated by IPSO save for the Guardian, the Financial Times and the Independent who declined to take part in it and established their own self-regulatory systems, while the Private Eye has never been part of a regulator.

160 Ibid.
162 R v Bitton [2019] EWCA Crim 1372.
163 Kay Goodall (n 152) 112; Hare (n 7) 529.
164 This is a statutory body created for the purpose of reviewing the law of England and Wales see Law Commission Act 1965, s 1.
IPSO has yet to apply for recognition by the PRP and does not intend to do so. The PRP suggests that it would be unlikely to fulfil the recognition criteria because it lacks independence from the publications it regulates. This is an issue that critics of IPSO have long identified and we shall shortly see how this appears to manifest in its decisions.

IPSO-regulated publications are expected to abide by the Editors’ Code of Practice (Editors Code). It regulates all of their output, including online and audio-visual material, where such content does not fall within the remit of the Broadcast regulator. Members of the public can make a complaint where they believe a publication has breached the code. IPSO will first try to amicably resolve the complaint. If this fails, it will investigate, adjudicate and determine whether the outlet breached the code. The only provision relevant to hate speech in the code is Clause 12 (i) and (ii), which provide that:

*The press must avoid prejudicial or pejorative reference to an individual's, race, colour, religion, sex, gender identity, sexual orientation or to any physical or mental illness or disability.*

And:

*Details of an individual's race, colour, religion, gender identity, sexual orientation, physical or mental illness or disability must be avoided unless genuinely relevant to the story.*

This is a replica of the discrimination clause under the PCC which has been criticised for only providing protection to individuals, not groups and therefore ‘ill-suited to countering the publication of articles containing general racist assumptions and

---

169 Ibid.
170 Ibid.
stereotypes’.\textsuperscript{173} It would therefore catch out a statement that ‘Ahmed Said is a Muslim and therefore a terrorist’ but not ‘All Muslims are terrorists’.\textsuperscript{174} IPSO justifies limiting the clause to individuals as necessary in order to ‘protect freedom of expression’, arguing that ‘it’s important in a free society that no group is considered above criticism, and that journalists are able to challenge the beliefs, practices, actions, opinions and traditions of different groups for the benefit of us all.’\textsuperscript{175} Publications are free to attack groups of people without breaching the clause, but they cannot single out an identifiable individual on account of a protected characteristic.

This was noted in the decision in \textit{Greer}, which was a complaint about a controversial article by the columnist Katie Hopkins titled ‘\textit{Rescue boats? I’d use gunships to stop migrants}’. The Article contained various inflammatory comments about migrants, including comparisons to ‘cockroaches’ and the ‘norovirus’ and suggesting that the government ought to borrow a leaf from the Australian response to refugees which she branded as ‘threaten[ing] them with violence until they bugger off, throwing cans of Castlemaine in an Aussie version of sharia stoning’.\textsuperscript{176} The committee confirmed its position regarding clause 12, noting that:

\textit{The Committee acknowledged the strength of feeling the column had aroused. It took this opportunity to note publicly that the terms of Clause 12 specifically prohibits prejudicial or pejorative reference to individuals; they do not restrict publications’ commentary on groups or categories of people. In this instance, the references under complaint were not to any identifiable individuals. As such, Clause 12 was not engaged.}\textsuperscript{177}

This limitation proved to be particularly controversial in \textit{Elgy} which related to a complaint about an article published in \textit{the Sun} titled ‘Now Philip Hammond is finally out he must shut the door behind him and take control over our laws, our trade and

\begin{itemize}
\item \textsuperscript{174} See for example Clare Balding v Sunday Times [2010] PCC.
\item \textsuperscript{176} \textit{Greer v The Sun} [2015] IPSO 02741-15.
\item \textsuperscript{177} Ibid.
\end{itemize}
especially immigration’. The article was an opinion column by Trevor Kavanagh—who interestingly was a member of IPSO’s Board at the time—claiming that there was:

_one unspoken fear, gagged by political correctness, which links Britain and the rest of Europe. The common denominator, almost unsayable until last week’s furore over Pakistani sex gangs, is Islam._

The article went on to suggest that ‘Muslims are a specific rather than a cultural problem’ and that the authorities had ‘long deliberately disregarded Muslim sex crimes – soon likely to be a racist offence in itself – including outrages such as female genital mutilation and “honour” killings’ and concluded by asking ‘what will we do about the Muslim Problem then?’

The Committee noted that the phrase ‘Muslim Problem’ could be interpreted as a ‘reference to the rhetoric preceding the holocaust’ and was ‘capable of causing serious offence’. However, it was aimed at Muslims as a group rather than a specific individual and therefore did not violate the code. It also clearly amounts to a DNS as it essentialises Muslims, labelling them as uniquely predisposed to committing sexual offences, and this is an undesirable trait that encourages and justifies prejudice.

Even where DNS are used against specific individuals IPSO’s adjudicative process is hamstrung by their superficial analysis of whether a publication has breached clause 12, usually confined to no more than one or two paragraphs. This is combined with a tendency to accept as fact the newspapers stated intentions for publishing a piece even where such explanations are unsupported by the contents. Like its predecessor, the PCC, its adjudications are ‘inconsistent and poorly reasoned’.

This was perhaps most evident in their decision in Manji, which related to a complaint filed by the Channel 4 journalist Fatima Manji about a column by Kelvin MacKenzie.

---

178 Elgy v The Sun [2016] IPSO 17562-17
179 Ibid.
180 Ibid.
181 Ibid.
182 Ibid.
184 Barendt and others (n 138) 39.
titled ‘why did Channel 4 have a presenter in a hijab fronting coverage of Muslim terror in Nice’ in which he asserted that he ‘could hardly believe his eyes’ that the complainant, ‘was not one of the regulars…but a young lady in a hijab’. He questioned whether it was ‘appropriate for her to be on camera when there had been yet another shocking slaughter by a Muslim’. He asked, ‘was it done to stick one in the eye of the ordinary viewer who looks at the hijab as a sign of slavery of Muslim women by a male-dominated and clearly violent religion?’ and further that ‘with all the major terrorist outrages in the world currently being carried out by Muslims, I think the rest of us are reasonably entitled to have concerns about what is beating in their religious hearts. Who was in the studio representing our fears?’

This complaint fell within clause 12 because it contained both generalised comments about Muslims and targeted comments that singled out the complainant. It is also arguable that the article made prejudicial and pejorative references to the complainant based on her religion. It certainly seemed to suggest that the complainant was either complicit in the terrorist attack by dint of being a Muslim or, at the very least, sympathetic towards the perpetrator’s motivations. It appeared to essentialise and blame Muslims collectively for the atrocity and suggest they have similar commitments as the attacker (‘what is beating in their collective hearts?’), labelling them with an undesirable trait by painting them as dangerous. It also seemed to classify the complainant as one of the collective blameworthy Muslims (‘who was in the studio expressing our fears?’), suggesting the presenter was not one of ‘us’ who had fears but was rather one of the Muslims who the ‘rest of us’ had good cause to fear. Despite this, IPSO dismissed the complaint. The committee noted that while the article referred to the complainant, it did so to trigger a legitimate debate of whether ‘newsreaders should be allowed to wear religious symbols’ and that the columnists was entitled to express his view on whether it was appropriate for a person wearing ‘Islamic dress’ to cover a ‘terrorist attack which had been carried out ostensibly in the name of Islam’.

The various flaws in IPSO’s adjudicative process manifested in this decision. The first is the lack of a detailed analysis of the substance of the allegation (the findings were

---

185 Manji v the Sun [2016] IPSO 05935-16
186 Ibid.
187 Ibid.
188 Ibid.
made in just two short paragraphs). The second is the acceptance of the columnist’s stated intentions at face value as just triggering a legitimate debate on the wearing of religious symbols. As Heawood notes, IPSO did not seem to understand that it is possible to ‘discuss religion without suggesting that every Muslim woman is a terrorist sympathiser’. 189 There is no obvious connection between triggering a public debate on the wearing of religious symbols and inferring that one who wears a hijab is somehow complicit in or sympathetic to a horrific terrorist attack. If such a connection does exist, it is certainly incumbent on IPSO to clearly set it out and certainly to spend more than two paragraphs analysing it. Otherwise, it is vulnerable to complaints of a lack of independence from regulated publications.

It is worth noting that even where a story does not breach clause 12, it can still amount to a violation of another part of the code, such as the accuracy clause which requires the press to ‘take care not to publish inaccurate, misleading or distorted information’. 190 The Muslim Council of Britain has pursued a particularly successful campaign under this provision obtaining corrections on at least 40 inaccurate stories about Muslims which would otherwise have been unaddressed under clause 12. 191

d. IMPRESS Standards Code

IMPRESS is the only Leveson-compliant regulator in the UK. It currently regulates around 100 publications which are mostly smaller local outlets compared to the large national newspapers with much greater circulation that are regulated by IPSO. 192 IMPRESS applies the Standards Code (‘Standards Code’) which borrows quite heavily from the Editors Code with a few notable additions.

It contains a discrimination clause that is a near carbon copy of clause 12 of the Editors Code prohibiting ‘prejudicial and pejorative’ references to an individual’s protected characteristics, with the only substantive difference being its more exhaustive list of protected characteristics. 193 The Standards Code also contains a separate ‘hate
speech clause’, which it described during the consultations for the code as ‘a significant departure from existing press regulation in England’.\textsuperscript{194} It provides that:

*Publishers must not incite hatred against any group on the basis of that group’s age, disability, mental health, gender reassignment or identity, marital or civil partnership status, pregnancy, race, religion, sex or sexual orientation or another characteristic that makes that group vulnerable to discrimination.*\textsuperscript{195}

As noted earlier, IMPRESS has mirrored the criminal law and has indicated in the code’s guidance that it will interpret this provision in line with the stirring up offences.\textsuperscript{196} Despite this, IMPRESS launched an investigation against one of its regulated publications, 5 Pillars UK, finding that they had breached the hate speech clause for publishing content which quite clearly did not fall foul of the stirring up offences.

5 Pillars UK is an online-only news website describing itself as a ‘Muslim community media platform’.\textsuperscript{197} The impugned content was contained in a podcast and fell within IMPRESS’s remit as similar to IPSO, they regulate all content produced by a regulated publication irrespective of the medium in which that content appears.\textsuperscript{198} During this podcast, the publication’s deputy editor claimed that ‘engaging in same-sex relations is a gross crime against Allah...this is something we cannot normalise within our communities or give legitimacy within our tradition’.\textsuperscript{199} IMPRESS found that this content breached the code as referring to homosexuality as a ‘gross crime’ may encourage and empower viewers to ‘take active steps to discriminate against LGBT Muslims who may or may not identify as members of the LGBT community, steps that might include threats of, or use of, physical violence’.\textsuperscript{200}

This content is homophobic though it is not DNS, or any stereotype for that matter. It is a theological claim about the status of homosexuality under Islam, one based on a


\textsuperscript{195} IMPRESS Standards Code 2017, cl 4.1


\textsuperscript{197} 5 Pillars [2020] 303/2020 IMPRESS 15.

\textsuperscript{198} ibid 17.

\textsuperscript{199} ibid.

\textsuperscript{200} ibid.
conservative interpretation of Islamic doctrine but by no means a fringe one. More importantly, as per IMPRESS’s guidance on interpreting the code, this content should not have breached it. The guidance explains:

The law defines incitement to racial hatred more broadly than incitement to hatred on grounds of religious belief and sexual orientation. This is because of the importance of allowing for free and vigorous debate on all matters concerning religious belief, observance, and practice and on matters of sexual morality and of respecting the freedom of traditional religious communities to, for example, voice their convictions that sex outside heterosexual marriage is sinful. IMPRESS’s approach to this Code provision will be similar.

This is clearly modelled on the freedom of expression clauses in the narrower stirring up offences. The similarities do not end there as the guidance also defines ‘threatening’ in the same terms as the criminal law noting that a threat that ‘certain groups will “burn in hell” should not be seen as constituting hate speech: threats must be of adverse consequences in this life, not an after-life.

The claim that homosexuality is a ‘gross-crime against Allah’, while bigoted, would certainly amount to the expression of religious views about the sinfulness of same-sex relationships and should not have breached the code. The publication was aggrieved and rightly pointed to the inconsistency between the high threshold for what constitutes hate speech under the Standards Code and guidance and the lower threshold IMPRESS applied in this case. IMPRESS tried to square this circle by distinguishing between a claim that homosexuality is sinful (which would be protected) and a claim that it is a crime (which would not be protected). This distinction is dubious at best and an unconvincing basis to delineate between protected and unprotected speech.

---

202 ‘Guidance on the IMPRESS Standards Code’ (n 196) 19.
203 ibid.
204 5 Pillars (n 201) 18.
205 ibid.
IMPRESS clearly felt that this content should not be protected under the hate speech clause even though the code and guidance, as framed, suggest it clearly ought to have been. It was forced to engage in a strained interpretation of the hate speech clause, leading to justified frustration on the part of the regulated publication, which as a result, threatened to withdraw from the regulator.206

IMPRESS tied its hands in this way because it decided to mirror the stirring up offences, a move informed by an adherence to a particular philosophical understanding of press freedom, an equivalence model that claims the press should have the same rights to publish hate speech as everyone else, and which is therefore uncomfortable with imposing any content restrictions that are broader than those imposed on the general public.207

e. Hate speech regulation in broadcasting

Discriminatory coverage of minority groups is not uniformly spread across the different mediums, and the literature set out in section 3 does not identify similar systemic issues in the broadcast media. By broadcast media here, I am referring specifically to that which publishes news content. These differences are to be expected because the broadcast media (radio and television) is subject to statutory regulation that contains stricter content standards. The justifications for these different regulatory approaches will be explored later.208 Here I will set out the content standards relevant to hate speech contained in the Broadcasting Code, noting that it regulates DNS through a broader hate speech clause and by engaging in a contextual analysis that is sensitive to the group-specific nature of stereotypes.

I. The Broadcasting code: A contextualised approach to hate speech

Broadcasters are required to comply with the Broadcasting Code, which contains various content standards.209 It also contains three provisions relevant to DNS. The first is Rule 2.3, which deals with ‘harm and offence’ and proscribes ‘discriminatory

---

207 See chapter 2.
208 See chapter 3.
209 For an overview of these standards see Barendt and others (n 138) ch 4.
treatment or language’ on the grounds of ‘age, disability, gender reassignment, pregnancy and maternity, race, religion or belief, sex and sexual orientation, and marriage and civil partnership’.

Two aspects of this are relevant. The first is the inclusion of a broader range of protected characteristics than those under either press code. Secondly, the guidance notes indicate that it covers stereotypes noting that using them in reference to an under-represented minority is offensive as it can create a ‘false impression of that minority’.²¹⁰ This clause applies to any stereotyping that may be found offensive by a general audience and is not concerned with whether they promote or justify prejudice. Therefore, it catches a much broader range of content than DNS.

The second and third provisions are clauses dealing with ‘hatred and abuse’. Rule 3.2 requires that material which ‘contains hate speech must not be included…except where it is justified by the context’. Hate speech is defined as including:

*all forms of expression which spread, incite, promote or justify hatred based on intolerance on the grounds of disability, ethnicity, gender, gender reassignment, nationality, race, religion, or sexual orientation.*

Before considering Rule 3.2 in greater detail, I will briefly mention Rule 3.3. It prohibits the inclusion of material ‘which contains abusive or derogatory treatment of individuals, groups, religions or communities…except where it is justified by the context’. This is similar to the discrimination clause found in the press codes, except it is expanded to include groups. In practice, Rules 3.2 and 3.3 are often read together and the determination that content breaches Rule 3.2 is often dispositive of both.

I will now turn back to Rule 3.2. This clause is broader than the stirring up offences in several respects. Firstly, it contains no qualifiers and refers to ‘all forms of expression’. This results in the coverage of hate speech that is not abusive, threatening or insulting and unlike the Standards Code, the guidance does not introduce these qualifiers through the back door. Secondly, it covers a broader range of effects, including speech

---

which promotes or justifies hatred rather than the more imminent standard of incitement.\textsuperscript{211} The result is that Ofcom regulates covert hate speech that is couched in more moderate language but which has the effect of justifying or encouraging prejudice.

However, this broader clause means that Ofcom will often have to adjudicate on more subtle forms of hate speech, and it is harder to determine whether such content has these effects. Overt hate speech is crude (‘Operation filthy Jew bitch’,\textsuperscript{212} ‘bomb a mosque day’\textsuperscript{213} or referring to a Nigerian as a ‘black monkey’\textsuperscript{214}), and it does not require a great deal of sophisticated linguistic analysis to determine the intention and effect of such speech. Less overt forms require more careful consideration and greater nuance to analyse and classify as hate speech. It requires Ofcom to engage in what Kay Goodall refers to as a ‘contextual meaning’ approach to incitement which considers the precise ways that hatred manifests against different groups.\textsuperscript{215}

To do so, one must first determine what kinds of stereotypes are used to encourage and promote prejudice against a specific group and thereafter examine whether the impugned content either explicitly or implicitly relies on them. Ofcom did exactly this in its decision in \textit{Islam Channel}, which related to a television programme alleging that the state of Israel was involved in distributing inaccurate copies of the Quran in the 1960s.\textsuperscript{216} The programme was broadcast in the Arabic language but contained English subtitles that stated:

\textit{the Jews … tried to distort [the Qur’an] when they printed hundreds of thousands of corrupted copies [in] 1961 … ‘and in their futile attempt to corrupt Al-Quran they [modern-day Jewish people] tried to destroy our beliefs and religion’, ‘by doing so the

\textsuperscript{213} ‘Hate Crime Laws a Consultation Paper’ (n 161) 16.
\textsuperscript{214} Ibid 76.
\textsuperscript{215} Kay Goodall (n 152) 97–99.
new Jews tried to do the same thing their ancestors [Jewish people in ancient times] did when they displaced words from (their) right places’ [of the Jewish Bible (Torah)].\textsuperscript{217}

The program also contained on-screen graphics which contained the following statement:

*Israel…continues with its poisonous acts with its attempt to change the meaning of the Qur’an. It [modern-day Israel] wants the obliteration of our beliefs and religion and in this way, it continues to practice what their forefathers [Jewish people in ancient times] had engaged in the past, particularly in their practice of changing the words [of the Jewish Bible (Torah)] in the past.*\textsuperscript{218}

Ofcom had to determine whether these statements amounted to hate speech against Jews. In doing so, it relied on the working definition of antisemitism by the International Holocaust Remembrance Alliance (IHRA) which provides that:

*Antisemitism is a certain perception of Jews, which may be expressed as hatred toward Jews. Rhetorical and physical manifestations of antisemitism are directed toward Jewish or non-Jewish individuals and/or their property, toward Jewish community institutions and religious facilities.*\textsuperscript{219}

The definition is accompanied by guidance which contains illustrations of various DNS that have been used to justify and promote hatred against Jews. Both the definition and accompanying guidance—which draws on the expertise of members of the IHRA Committee on antisemitism and how it manifests—are regarded as authoritative and have been officially adopted by the governments of 25 countries, including the UK.\textsuperscript{220}

It is worth noting that the aspects of ‘new antisemitism’ in the IHRA definition are not universally accepted and some have suggested it has been used to stifle debate about

\textsuperscript{217} ibid 17.
\textsuperscript{218} ibid 23.
the policies of Israeli and target supporters on the Palestinian side of the Israeli-Palestinian conflict.\(^221\)

Ofcom noted that the impugned content contained various Antisemitic DNS referred to in the IHRA guidance. Firstly, it ascribed a ‘perpetually negative characteristic to Jewish people’, accusing them of being ‘tyrannical’ and having an ‘evil mind’ and engaging in a conspiracy to destroy Islam in the past and today. This is described in the IHRA guidance as ‘Accusing Jews as a people of being responsible for real or imagined wrongdoing committed by a single Jewish person or group, or even for acts committed by non-Jews’, a classic antisemitic DNS.\(^222\)

Secondly, the broadcast and subtitles interchangeably use the term Israel and Jews when describing the culprits responsible for this alleged plot of distributing inaccurate copies of the Quran. This conflation of the actions of the state of Israel and Jewish people is another Antisemitic DNS set out in the IHRA guidance. By holding Jews collectively responsible for this alleged attempt at distorting the Quran, the program justifies why Muslims should hate the state of Israel and Jews, which the speaker treats as synonymous. Ofcom found that the broadcast ‘had the potential to promote, encourage and incite such intolerance among viewers’.\(^223\) Ofcom also relied on the IHRA definition and guidance to a similar effect in another decision in *Malik Live*.\(^224\)

There are two reasons that Ofcom successfully regulates DNS. Firstly, it recognises what constitutes hate speech should not be made in a vacuum but must be contextualised by considering what Allen terms the ‘historical, social, cultural, and political processes’ informing how that phenomenon manifests against specific groups.\(^225\) This means that definitions and guidance such as that provided by the IHRA are indispensable when trying to regulate DNS. Secondly, unlike IPSO and IMPRESS


\(^{222}\) ‘Working Definition of Antisemitism’ (n 219).

\(^{223}\) ‘Ofcom’s Broadcast and On Demand Bulletin’ 24.


which confine their analysis of the content to two or three paragraphs, Ofcom carefully scrutinises the content in some detail, with the analysis often spanning several pages. This is not to suggest that Ofcom’s approach is perfect; far from it, and its weaknesses are considered in chapter 9. However, it is far better equipped to deal with discriminatory content published about groups than the press regulators.

7. Conclusion

This chapter has explained that the discriminatory reporting of ethnic minorities in the UK press has long been of concern to scholars and human rights bodies. It also noted that Leveson recommended that a new press regulator ought to be able to regulate such content. It has also been identified that the concerns raised by these bodies and by Leveson are about a form of covert hate speech, what I have termed as DNS, which essentialise and assign undesirable traits to minorities.

I have examined the two major press codes noting that neither regulates this content. IPSO, the body which regulates most newspapers simply doesn’t regulate hate speech against groups. IMPRESS, on the other hand, does so but mirrors the stirring up offences in the Standards Code which is limited to overt hate speech. I have noted that in practice, it appears to have taken a broader interpretation which appears to be inconsistent with the Standards Code’s guidance.

This stands in stark contrast to the Broadcast sector, where DNS are regulated through a more broadly formulated hate speech clause that catches covert hate and a contextualised approach that is willing to rely on group-specific guidance to determine what constitutes hate speech.

The purpose of this thesis is to examine whether similar restrictions should be imposed on the press. In the next chapter, I frame both the questions that will need to be answered to justify this and the methodology I will use to answer them.
Chapter 2: Methodology and Research Questions

1. Introduction

I have shown that the discriminatory content appearing in some press outlets is a form of covert hate speech, what I have termed dangerous negative stereotypes and that this issue remains unaddressed in the post-Leveson regulatory framework.

I have also noted that the broadcast media is subject to more stringent content standards which include a broader hate speech clause that catches some DNS. This thesis explores whether applying similar restrictions to the press is both desirable and compatible with press freedom. It does so by exploring and answering the following research questions:

1. What kind of hate speech does the press publish, and should it be restricted?
2. What is the relationship between freedom of expression and press freedom, and how can these principles be reconciled with restrictions on covert hate speech?
3. Does this hate speech cause harm, and if so, how?
4. What form should effective press-specific hate speech regulation take? To whom should it apply? And where would it fit in the current regulatory framework?

This chapter will expound on the methodology that will be used to answer these questions and frame the issues that will be considered in the rest of this thesis. The first question has partly been answered as I have already identified the kind of hate speech that the press publishes. Section 2 focuses on the second question. I will identify what the prominent theories of press freedom are and how they conceptualise the relationship between freedom of expression and press freedom. It will be suggested that a good answer to the research questions must account for both the libertarian theory preferred by the industry and the social responsibility theories preferred by scholars.

Section 3 deals with the third question. I will explain my reliance on the coverage/protection distinction to determine the scope of each of these principles. I will also set out how the limits to each principle will be informed by a balancing approach and a harm-prevention framework. Thereafter I identify the conditions under
which broader regulation is justifiable under these theories and frame the issues that will be considered in the chapters that follow. The fourth question is explored and answered in chapter 9.

2. Theories of press freedom

The relationship between press freedom and freedom of expression has been conceptualised in different ways. These are set out by Fred S. Siebert in *Four Theories of the Press*, which the media historian James Curran describes as the ‘bible of comparative media studies’. The core of Siebert’s thesis is that the press’s functions are informed by the specific social and political structures in which they are situated. The four theories of the press were predicated on the divergent political systems (at the time of writing); the libertarian, social responsibility, authoritarian and soviet-communist theories. My analysis is limited to the libertarian and social responsibility theories which are of direct relevance to liberal democracies. I also draw on Helen Fenwick and Gavin Phillipson’s taxonomy of press freedom as well as a recent contribution by Paul Wragg.

a. Libertarian theory and the equivalence model

According to Siebert, the authoritarian theory was replaced by the libertarian theory in the post-enlightenment period with the change informed by the rise of liberal democracy and the ascendancy of ‘laissez-faire economics’. What Siebert refers to as ‘libertarianism’ is a particular school of thought within broader libertarian political philosophy, distinguishable from socialist or anarchist conceptions. It would perhaps best be described as ‘right-libertarianism’, which, as Wragg notes, sees the market, nor the state as the ‘only means of realising press freedom’.

Drawing on John Locke, John Milton and John Stuart Mill, the premise of the libertarian theory is largely (though not exclusively) centred on the ‘marketplace of ideas justification’ for freedom of expression. It is explored in a later chapter, but its central

---

228 Ibid 2–3.
229 Wragg, *A Free and Regulated Press* (n 18) ix.
premises are worth mentioning here as they are of relevance. It assumes that men are rational beings who are capable of distinguishing right from wrong and that truth will win out over falsity in a ‘free and open encounter in the marketplace of ideas.’ \(^{230}\)

Libertarian theory sees the media as a tool to aid in truth-seeking by ‘informing and entertaining’ its audience who could be trusted to sift the good ideas from the bad in a ‘self-righting’ process. \(^{231}\)

How then does this theory conceive of the relationship between freedom of expression and press freedom? There is some divergence on this point. Some libertarian theorists see them as synonymous. That press freedom means the same thing as freedom of expression. Eugene Volokh argues that this is the nature of press freedom under the First Amendment to the US Constitution (‘First Amendment’), which protects press freedom as a ‘technology’. \(^{232}\) Put simply, the law protects the right to exercise one’s free speech rights using technology such as the printing press and its ‘modern equivalents’. \(^{233}\) This view does not recognise any special duties or responsibilities that the press owes to the public or, as Siebert puts it, ‘a newspaper is a private enterprise owing nothing to the public’. \(^{234}\)

Other libertarians argue that press freedom is distinct from freedom of expression. Wragg notes many of its adherents accept the press have an instrumental role in exposing abuses of power and are a vital part of the system of checks and balances. \(^{235}\) However, despite their disagreements as to the nature of this relationship, they arrive at the same conclusions regarding their scope. This is because libertarians in the second camp contend that because the press operates as a check on government, it cannot accept special rights, privileges, or responsibilities which they argue could only lead to ‘domination’. \(^{236}\) Accepting special privileges means also taking on special responsibilities, which can then be enforced on the press by the state. Baker argues that this would be allowing ‘the fox to watch the door to the hen house’. \(^{237}\)

---

\(^{230}\) Siebert and others (n 227) 3, 45 and 51.

\(^{231}\) Ibid 3, 45, 51.

\(^{232}\) Ibid 73.


\(^{234}\) Siebert and others (n 227) 73.


\(^{236}\) Siebert and others (n 227) 52.

points out that the UK media have been reluctant to accept special privileges on this basis and prefer press freedom to have the same scope as freedom of expression.\textsuperscript{238}

Rather than accept these special privileges and responsibilities, the metaphorical marketplace of ideas in the press is to be regulated and supported by a literal economic market. Papers will sell, the ones with the most public support and, therefore, the best ideas (as man is a rational being) would rise to the top.\textsuperscript{239} Free enterprise, not government, would regulate quality and ‘Adam Smith’s free hand and Milton’s self-righting process would set things right’.\textsuperscript{240}

Libertarians then, either through viewing press freedom and freedom of expression as the same thing or through accepting they are different but, for pragmatic reasons, believe they should be treated the same, arrive at the same conclusion. The scope of press freedom ought to be the same as freedom of expression. This has led Fenwick and Phillipson to refer to this as the ‘equivalence model’.\textsuperscript{241}

What implication does this equivalence model have on the scope of hate speech laws? Superficially, if the press have equivalent speech rights to members of the public, specifically those members who wish to engage in hate speech (‘lone hate speakers’), it would suggest that no greater restrictions on hate speech can be imposed on them.

b. Social responsibility theories and the variable geometry model

The second model comprises of social responsibility theories (‘SRT’)\textsuperscript{242} which Siebert argues arose out of doubts in the social sciences about the rationality of man.\textsuperscript{243} SRT’s questioning of the central premises of the libertarian theory is borne out of developments in the press during the post-war period. Siebert identifies several issues existing at the time, which are striking only in their familiarity. Publishing inaccuracies, owners influencing editorial content, control of the media by the wealthy, a focus on sensationalism over substance and a penchant for invading privacy.\textsuperscript{244} SRT saw this

\textsuperscript{238} Rowbottom, \textit{Media Law} (n 42).
\textsuperscript{239} Siebert and others (n 227) 52.
\textsuperscript{240} ibid 76–77.
\textsuperscript{241} Fenwick and Phillipson (n 20) 20.
\textsuperscript{242} I borrow this abbreviation from Wragg, \textit{A Free and Regulated Press} (n 18).
\textsuperscript{243} Siebert and others (n 227) 99–100.
\textsuperscript{244} ibid 79,80.
oversized influence of the media and its capacity to do harm and doubted whether the rationality of man or the markets could reign in press excesses. The essence of SRT is reflected in the following statement by Baroness Onora O’Neill:

*Communication of the powerful can shape and influence, improve and damage others’ lives, and in democracies we have long since taken steps to regulate the communication of most powerful organisations.*

Moreover, unlike libertarian theory which focuses on truth-seeking, SRT are principally concerned with ‘promoting a harmonious and fruitful society’ and see the functions of the media through this lens. Siebert sets out some of these functions as ‘providing information, discussion and debate on public affairs’, ‘enlightening the public’ and ‘serving as a watchdog against government’. SRT also acknowledge the role of the state in creating an environment that allows the press to fulfil their functions by granting them special privileges. In an ideal world, the press would also voluntarily abide by the duties and responsibilities these functions require. However, if they fail to do so, SRT are more comfortable than libertarian theory with state intervention which holds them to these duties and responsibilities.

SRT hold great currency and have informed the discussions of the various Royal Commissions and press inquiries formed to investigate press standards. They are also preferred in the UK academic literature appearing in the work of Eric Barendt, Rachael Craufurd-Smith and Damian Tambini, Fenwick and Phillipson, Gibbons, Oster, O’Neill, and Rowbottom, among others, and form the basis of arguments for enacting and enforcing certain standards on the press such as requirements to aim for accuracy, as well as structural regulation to ensure pluralism and diversification of ownership.

---

246 Siebert and others (n 227) 102.
247 ibid 74.
248 ibid.
249 The British Royal Commission on the Press, Cmd. 6810 (1977), para 2.3; Leveson (n 2) 56–63.
How then does SRT conceive of the relationship between freedom of expression and press freedom? Theorists in this camp generally regard press freedom as a distinct concept from freedom of expression, one which is protected only to the extent that the press fulfil their essential functions. What SRT are divided on, however, is whether and to what extent the press has recourse to both arguments for press freedom and freedom of expression.

These divisions can broadly be categorised into what Fenwick and Phillipson describe as the ‘differentiated privileges’ and the ‘variable geometry’ models. Oster is one of the proponents of the former which sees press freedom as ‘special vis-à-vis freedom of expression’ in both scope and intensity of protection.\(^{251}\) Scope means that press freedom entitles the media to privileges which other individuals do not possess, and greater intensity means that restrictions on press freedom should be subject to greater scrutiny than restrictions on freedom of expression.\(^{252}\) These special privileges include the protection of sources and journalistic material as well as special rights of access.\(^{253}\)

Under this model, claims for this heightened protection and special privileges are justified only when the press perform SRT functions, i.e. when they act as watchdogs, but not when they unjustifiably intrude into an individual’s privacy.\(^{254}\) However, this model only subjects claims for press freedom (heightened protection and special privileges) to such scrutiny. This means that although the press may be denied such protection and privileges when they fail to fulfil these democratic functions, they would still ‘enjoy the general freedom of expression protection under the law’.\(^{255}\)

The variable geometry model, on the other hand, subjects all press speech claims to such scrutiny. Claims for the same speech rights as members of the public are only justified if they are supported by free speech rationales, and their claims for special privileges are warranted only when they fulfil their functions.\(^{256}\) This model is preferred in my analysis because it is more thorough and is in line with the consensus in media

---


\(^{251}\) Oster (n 250) 59.

\(^{252}\) Ibid 59–61.

\(^{253}\) Oster (n 24) ch 4.

\(^{254}\) Wragg, *A Free and Regulated Press* (n 18) ch 7.

\(^{255}\) Oster (n 250) 61.

\(^{256}\) Fenwick and Phillipson (n 20) 26–27.
law scholarship which is sceptical about whether mass media institutions can rely on the same justifications for free speech as lone speakers.\textsuperscript{257}

c. Bridging the divide

Wragg identifies what he terms the ‘ideological divide’ over press regulation, arguing that the deadlock in debates around press regulation is informed by the opposing groups having contrasting views of press freedom.\textsuperscript{258} On one side of the debate is most of the press which adopts the libertarian view and uses it to resist any attempts at enforcing any duties and responsibilities through statutory regulation. On the other side are SRT reflected in the various commissions of inquiry and the academic literature which insists that these duties and responsibilities can coercively be enforced by the state.

One way of resolving this deadlock is to accept the scholarly consensus around SRT and simply inquire as to whether broader hate speech regulation is compatible with them. One could even argue that current press regulation supports and accepts SRT claims. For example, if the press is so committed to the libertarian view, what is the purpose of the Editors and/or Standards Code? These would surely be superfluous if the press really believed in equivalent treatment. By imposing on themselves various duties and responsibilities which are not required of other publishers, such as requirements to avoid inaccuracies\textsuperscript{259} and even broader restrictions on hate speech directed against individuals than can be found under the general law,\textsuperscript{260} the press arguably accept SRT claims that they owe certain duties and responsibilities. The press codes are how they are enforced. Surely then, all that is required is merely a better-formulated provision that would proscribe the perpetuation of DNS against groups coupled with more robust enforcement mechanisms.

A look at the history of press regulation tells a different story. It is not one in which the press embraced SRT and welcomed the idea of regulation. They were instead given a stark choice, self-regulate, or you will be regulated by statute. The first press code

\textsuperscript{257} See Chapter 7 and generally Gavin Phillipson, ‘Leveson, the Public Interest and Press Freedom’ (2013) 5 Journal of Media Law 220.
\textsuperscript{258} Wragg, A Free and Regulated Press (n 18) ch 2.
\textsuperscript{259} IPSO Editors Code 2021, cl 1; IMPRESS Standards Code 2017, cl 1
\textsuperscript{260} IPSO Editors Code 2021, cl 12; IMPRESS Standards Code 2017, cl 4.1
was promulgated because of a fear of imminent statutory regulation in light of the (then ongoing) 1989 Calcutt inquiry into rampant abuses by the tabloid press.\textsuperscript{261} The political climate was reflected in the Home Secretary’s statement that ‘this is positively the last chance for the industry to establish an effective non-statutory system of regulation.’\textsuperscript{262} Similarly, the current largest regulator, IPSO, was formed after the Leveson Inquiry, with Lord Leveson infamously noting, ‘I cannot, and will not, recommend another last chance saloon for the press.’\textsuperscript{263}

The taking on of additional obligations under the codes does not undermine the libertarian theory as these are merely contractual arrangements and can be discarded by press outlets terminating their membership with the respective regulators.\textsuperscript{264} Regulated publications also vote on the contents of the Editors’ code.\textsuperscript{265} These codes simply reflect the extent of the responsibilities the press are comfortable voluntarily assuming. It is easy for the press to comply with hate speech constraints that prevent them from making pejorative references about an individual’s protected characteristics when the primary form of hate speech they engage in is the publication of DNS against groups. To them, self-regulation is simply the lesser of two evils. It is pragmatism, not principles which informs the current framework.

Why should this matter? After all, the prevalent view is that the equivalence model is misguided and should be rejected. This is in fact a view that this thesis shares. Specifically, in the context of hate speech, it is submitted that the most persuasive accounts for a ‘right’ to engage in hate speech are non-instrumentalist speaker-based accounts which do not apply to or are weaker in the case of the press.\textsuperscript{266} If this is the case, surely efforts are better spent convincing the press to embrace SRT failing which legislators ought to put in place a statutory framework proscribing DNS.\textsuperscript{267}

\begin{itemize}
\item \textsuperscript{261} O’Malley and Soley (n 139) 88.
\item \textsuperscript{262} Wragg, A Free and Regulated Press (n 18) 265.
\item \textsuperscript{263} Leveson (n 2) 1757.
\item \textsuperscript{264} Publications that sign up to regulation by IPSO and IMPRESS both have the right to withdraw from the regulators by written notice. See IPSO Scheme Membership Agreement 2019 https://www.ipso.co.uk/media/1813/ipso-scheme-membership-agreement-2019-v-sep19.pdf and IMPRESS Regulatory Scheme Agreement https://pressrecognitionpanel.org.uk/wp-content/uploads/2016/05/IMPRESS-Regulatory-Scheme-Agreement-revised.pdf
\item \textsuperscript{265} Ibid cl 7.
\item \textsuperscript{266} See chapter 7.
\item \textsuperscript{267} O’Neill uses such an approach to argue for the imposition of certain accuracy standards Onora O’Neill, Regulating for Communication (2012); O’Neill (n 250).
\end{itemize}
The problem with this approach is that it does not take seriously the importance of seeking common ground with the press. Leveson was at pains to point out that press regulation must be able to secure the confidence and support of the entire industry.\footnote{Leveson (n 2) 1758, 1771,1782,1794.}

Any attempt at introducing restrictions on the press would encounter significant resistance from the industry. But arguments based on theoretical claims that they reject are likely to be the least persuasive. On the other hand, arguments that can be squared with (or at least address) their preferred equivalence model are more likely to be accepted. Adopting such an approach also lends greater legitimacy to coercively imposing these standards if the press refuse to apply them voluntarily.

To bridge this divide. This thesis will examine whether broader regulation of hate speech in the press can be reconciled with both the libertarian (equivalence model) and SRT (variable geometry model) theories of press freedom. I will first set out the methodology that will be used in this examination and thereafter frame the issues to be considered in subsequent chapters.

3. Methodology

This thesis is primarily a philosophical inquiry. It examines the theories underpinning freedom of expression and press freedom and what they inform us about the kinds of hate speech restrictions which ought to be imposed on the press. However, it is an inquiry grounded in a particular political context, one which accepts the premise that viewpoint punitive restrictions on hate speech of the kind that can be found in most liberal democracies, including Britain, are justifiable. This means that absolutist or near absolutist conceptions of press freedom or freedom of expression which argue for entirely de-regulating hate speech are not given serious consideration simply because they do not fit with European norms on hate speech.\footnote{Frederick Schauer, ‘Freedom of Expression Adjudication in Europe and the United States: A Case Study in Comparative Constitutional Architecture’ in Georg Nolte (ed), European and US Constitutionalism (Cambridge University Press 2005); Eric Heinze, ‘Wild-West Cowboys versus Cheese-Eating Surrender Monkeys: Some Problems in Comparative Approaches to Hate Speech’, Extreme Speech and Democracy (Oxford University Press 2009).} I also examine the positive law only to the extent necessary to determine whether my specific proposals to regulate
DNS would pass muster under Article 10 of the European Convention on Human Rights (‘Article 10’).270

My examination will be informed by a balancing approach which relies on Frederick Schauer’s distinction between coverage and protection. Before setting out the nature of this distinction and how it will be used in my analysis, I will briefly address questions of comparative method.

a. A note on comparative method

Much of the literature on free speech, press freedom and hate speech is written by American scholars who are writing either explicitly (or implicitly) from the perspective of the First Amendment. A theoretical examination of these issues would be incomplete without taking this body of literature into account. Moreover, the coverage/protection distinction which I rely on is developed by a First Amendment scholar. It is therefore pertinent to deal with questions of comparative methodology. This is because relying on these theories and the coverage/protection distinction as an analytical tool brings about the spectre of what Heinze aptly describes as the ‘Euro-American’ divide on hate speech.271

It is argued that America differs so considerably from other liberal democracies on questions of free speech and hate speech that cross-cultural comparisons of legal norms and approaches are riddled with pitfalls and traps. These differences are attributed to history,272 politics,273 culture274 and constitutional architecture.275 First Amendment doctrine utilises a different analytical method notable for its careful categorisation and the application of a distinctively high degree of scrutiny to any

270 See chapter 9.
271 Hare (n 60) 183. While Heinze uses this term he also—rightly—notes that various other countries outside of Europe adopt similar hate speech bans ibid 186.
272 Samuel Walker, Hate Speech: The History of an American Controversy (University of Nebraska Press 1994) 12–15; Hare (n 60) 531.
275 Walker (n 272) 12–15.
viewpoint punitive restrictions on speech. European countries, on the other hand, are more content to balance rights against countervailing interests through a proportionality analysis.

My analysis will be cognizant of this ‘Euro-American’ divide. The extent to which it is relevant depends on what aspects of the American scholarship I am examining. The differences are far less relevant when examining the philosophical arguments in support of either a principle of free speech or of press freedom or the harms caused by hate speech. These are questions that transcend the specific historical, cultural, and constitutional circumstances of either side of the Euro-American divide. Similarly, the arguments for why hate speech is harmful are salient in both contexts. The Law Commission’s review of hate speech laws is infused with harm theories drawn from American scholarship.

The Euro-American divide appears to be much more relevant in my reliance on the coverage/protection distinction. This is a distinction that was popularised by Frederick Schauer, a prominent American legal scholar. I develop the details of the approach below and will not repeat it here. I will however note that the use of this method is appropriate for two reasons. Firstly, it is entirely possible to divorce the distinction from any method of constitutional interpretation. Schauer in fact presents it as an analytical tool that is useful to explore free speech as a political principle. Secondly, even if Schauer has often used it as a technique of First Amendment analysis, it is not a method that is peculiar to it. He notes, for example, that it can be found in any approach to rights adjudication and that it is incorporated in the proportionality

---

278 ‘Hate Crime Laws a Consultation Paper’ (n 161) 65–72.
281 See generally Schauer, ‘Speech and Speech--Obscenity and Obscenity: An Exercise in the Interpretation of Constitutional Language’ (n 279).
analysis—which is the method of adjudicating interferences with rights under the ECHR.\textsuperscript{282} The inclusion of this scholarship and the reliance on the coverage/protection distinction does not import the areas in which the Euro-American divide matters.

b. Coverage

Schauer argues that understanding the scope and limits of free speech requires determining questions of coverage (whether an activity counts as speech) and protection (the degree of protection which the activity ought to be afforded).\textsuperscript{283} He posits that neither coverage nor protection can be understood without reference to the theories underlying a free speech principle. I will first set out the consequences of recognising such a principle and thereafter explain the reliance on theory to determine its coverage, protection and ultimately, the circumstances under which it can justifiably be restricted.

The appropriate limits of state power over the individual form the substance of much debate in political philosophy. Some argue for a minimalist state serving the role of a ‘nightwatchman’, which only protects a citizen’s rights from infringement by others.\textsuperscript{284} Others envision a larger or more positive role for the state.\textsuperscript{285} All conceptions of the role of the state will have within them rules that determine when it can legitimately interfere with another’s conduct, what Schauer terms the ‘normal standard of justification’ (‘baseline standard’).\textsuperscript{286}

While there are significant disagreements about the best argument for a free speech principle, there is broad—though not universal—agreement that such a principle does exist.\textsuperscript{287} The effect of the principle is to create an exception to the baseline standard by requiring states to provide stronger reasons for restricting speech than they must provide for restrictions on other domains of human conduct. As Mary Kate McGowan

\textsuperscript{283} Schauer, Free Speech (n 279); Schauer, ‘What Is Speech?’ (n 280).
\textsuperscript{284} Robert Nozick, Anarchy, State, and Utopia (Basic Books 1974).
\textsuperscript{286} Schauer, Free Speech (n 279) 7.
succinctly puts it, it ‘takes more to justify regulating speech than actions that are not speech’. As an example, Barendt points out that a state may limit the construction of buildings on purely aesthetic grounds, but such a rationale would not be sufficient to justify regulating speech.

Defining something as ‘speech’ carves out special protection for an activity by granting it greater immunity from state interference. One way of determining what speech means is by relying on its ordinary usage as ‘verbal and written utterances’. This would be incredibly under-inclusive in some ways and therefore leave out a plethora of unconventional acts which would fall within a free speech principle, such as a sit-in protest or flag-burning. In other ways, it would be incredibly over-inclusive and include utterances which would never conceivably fall within a free speech principle, such as making contracts, perjury or extortion.

These problems are lessened by using the term ‘freedom of expression’ which appears in Article 10. The term ‘expression’ in its ordinary sense would suggest a wider coverage encompassing not just verbal or written utterances but arts and dramatic performances. However, the issues are not entirely resolved. The term expression is also overinclusive as there are many actions that ‘manifest the inner feelings of the actor’ which are not covered by a free speech principle, such as punching someone, selling heroin or hiring a hitman. Freedom of speech or expression is too vague, ‘open-ended and morally loaded’ and cannot be properly understood without reference to the values that underly it.

This also applies to a principle of press freedom. This principle has a similar effect by requiring the state to provide stronger justifications than the baseline standard when

288 Mary Kate McGowan, Just Words: On Speech and Hidden Harm (Oxford University Press 2019) 158.
291 Greenawalt (n 287) 3.
292 Barendt, Freedom of Speech (n 6) 5.
293 Schauer, Free Speech (n 279) 13; Barendt, ‘What Is the Harm of Hate Speech?’ (n 289) 549.
295 Schauer, Free Speech (n 279) 89.
296 ibid 91; Barendt, Freedom of Speech (n 6) 4–6.
297 Oster refers to this as a Media Freedom principle Oster (n 24) 2.
interfering with certain activities in newsgathering and publishing. As with a free speech principle, some activities are not covered by a principle of press freedom. It is not implicated when a media outlet is required to comply with building regulations and obtain planning permission to renovate its offices.298 Similarly, hacking people’s phones would incur liability under criminal law and the tort of misuse of private information.299 However, these laws do not infringe on press freedom because the press is not protected in order to carry out such activities.300 Rather, this principle only covers those activities fulfilling the various democratic functions assigned to the press.301 Questions of coverage cannot, therefore, be determined without recourse to the justificatory theories underpinning either principle.

c. Protection

Once it is ascertained which activities are covered, the extent of protection afforded to them by the principle must be determined. One way of conceiving of this is that the question of coverage is dispositive of the question of protection (coverage=protection).302 However, this approach is too simplistic as the protection limb is distinct and comprises different considerations. To explain protection, Schauer draws a useful analogy to a suit of armour which I borrow with necessary modifications.303

A principle of free speech or press freedom is a suit of armour. An activity that is covered by the principle is clad in armour. But all armour is not built the same and will offer some degree of protection from some ammunition but not others.304 Some armour is sturdy enough to withstand rocks, but no armour will protect against artillery fire. The protection the armour affords an activity is determined by two factors. The first is the armour’s ballistic properties. This is a question of constitutional architecture.

A free speech principle does not necessitate absolute protection for covered speech but merely that the state must provide a greater justification to restrict speech than

298 ibid 131–132.
299 Gulati & Ors v MGN Ltd [2015] EWHC 1482 (Ch)
300 Oster (n 24) 37; Wragg, A Free and Regulated Press (n 18) 6–7.
301 These functions are examined in Chapter 8
302 Schauer, Free Speech (n 279) 90.
304 the ammunition being the justification for regulating the covered activities.
non-speech conduct. This justification can be significantly or only slightly more than the baseline standard.\textsuperscript{305} We would say that a legal system which requires a much greater justification than the baseline standard applies a stronger free speech principle than a state requiring only a slightly greater justification. But the latter still recognises a principle. Armour made of steel is stronger and can protect one against more projectiles than armour made of iron, but the latter is still better than no armour.

The US recognises a strong free speech principle through a combination of a strong form of judicial review and stringent legal tests on restrictions of speech covered by the First Amendment, e.g. the Brandenburg doctrine, which requires the state to show the speech would produce ‘imminent lawless action’.\textsuperscript{306} On the other hand, speech not covered by the First Amendment can be restricted on the very minimal ‘rational basis’ standard.\textsuperscript{307}

The UK recognises a weaker principle through a combination of a weak form of judicial review where courts have no power to strike down legislation\textsuperscript{308} and a proportionality analysis which grants legislators greater deference to determine the appropriate scope of restrictions of speech covered by Article 10.\textsuperscript{309} However, it still provides some protection relative to activities not covered by the right. Restrictions on uncovered activities are usually not amenable to any judicial review, and other public law claims often require demonstrating the restriction was illegal, procedurally improper, or irrational.\textsuperscript{310}

The second factor determining protection is whether speech is covered by the armour’s strongest or weakest points. This is a normative question, and it can only be determined by reference to the values underpinning the principle. Speech at the core

\textsuperscript{305} Schauer, Free Speech (n 279) 9.


\textsuperscript{307} Ferguson v Skrupa, 372 US 726 (1963)

\textsuperscript{308} They can instead declare that primary legislation is incompatible with the convention. This declaration empowers (but does not compel) the relevant minister to make a remedial order to remove the incompatibility see Human Rights Act 1998, s 10; Aileen Kavanagh, ‘What’s so Weak about “Weak-Form Review”? The Case of the UK Human Rights Act 1998’ (2015) 13 International Journal of Constitutional Law 1008.


of the principle, where the armour is toughest, ought to benefit from greater protection than that which is covered but lies at the margins.\textsuperscript{311} If the only sound justification for a free speech principle is to enable democratic self-government, protesting a proposed tax law would lie at the core of the principle and ought to have greater protection than artistic speech, even if the latter is also covered by the right.\textsuperscript{312} A practical example of such an approach is the hierarchy of protected speech in the ECtHR’s jurisprudence, with ‘high-value’ speech, such as political speech\textsuperscript{313} benefiting from the greatest protection, followed by artistic speech\textsuperscript{314} and thereafter ‘low-value’ speech, such as commercial advertising.\textsuperscript{315} Restrictions on high-value speech are subjected to a higher degree of scrutiny with states required to provide ‘very strong reasons’ for its restriction.\textsuperscript{316} Restrictions on low-value speech only require showing that there are ‘reasonable grounds’ justifying the restriction.\textsuperscript{317} The same analysis applies to a principle of press freedom. Though the principle may cover both celebrity gossip and revealing wrongdoing by the government, only the latter would benefit from the strongest protection.

We can then roughly sketch out how activities can be classified under the coverage/protection distinction. First are those which are not covered by either principle (‘no-value’), those which are covered but are at the periphery and benefit from the weakest protection (‘low-value’) and those which benefit from the strongest protection (‘high-value’).

\begin{footnotesize}
\begin{enumerate}
\item Harris and others (n 313); Maya Hertig Randall, ‘Commercial Speech under the European Convention on Human Rights: Subordinate or Equal?’ (2006) 6 Human Rights Law Review 53.
\item X and Church of Scientology v Sweden (1979) Application number 7805/77; Casado Coca v. Spain [1994] Application no. 15450/89
\item Stefan Sottiaux, “Bad Tendencies” in the ECtHR’s “Hate Speech” Jurisprudence’ (2011) 7 European Constitutional Law Review 40, 45.
\end{enumerate}
\end{footnotesize}
I. Harm prevention and balancing

It may be that the justifications for these principles simply do not support or apply to hate speech, and it is no-value under either.\textsuperscript{318} If this were the case, hate speech would be regulable on the baseline standard, and legislators should have almost unfettered discretion to determine what limits they want to place on it. In practical terms, this is reflected in the ECtHR’s exclusion of some forms of hate speech from the coverage of Article 10.\textsuperscript{319}

However, in other cases, it may well be covered and protected by these principles. For example, if free speech is important because it respects the speaker’s formal autonomy,\textsuperscript{320} it may well include a right for the speaker to disclose their values, including hateful ones.\textsuperscript{321} There may also be occasions on which the press may need to publish hate speech in order to perform its functions as a watchdog.\textsuperscript{322} Hate speech may therefore be covered and amount to either low or high-value speech under the respective principles. In these cases, these principles may come into conflict and must be balanced against other values and interests. Before setting out what this balancing entails, I will first explain why harm prevention is the appropriate interest to be balanced against freedom of expression and press freedom.

II. Liberalism and a principle of harm-prevention

Many scholars have presented various arguments to justify regulating hate speech. The vast array of rationales makes selecting a candidate justification for regulation difficult.\textsuperscript{323} One way of addressing this is to adopt a value pluralistic approach which recognises that ‘very different kinds of moral arguments can be made for and against hate speech law’, settling on the ones that are most persuasive and applying them to press hate speech.\textsuperscript{324} This certainly has its benefits, but its diversity is also the source of its greatest weakness. As Robert Mark Simpson writes, the vast spectrum of

\textsuperscript{318} Waldron, The Harm in Hate Speech (n 14) 147.
\textsuperscript{320} See chapter 7
\textsuperscript{321} Greenawalt (n 287) 300–301.
\textsuperscript{322} See chapter 8
\textsuperscript{323} Simpson (n 9) 704–705.
\textsuperscript{324} Brown, Hate Speech Law (n 55) 51.
reasons leads to perennial debates as any account ‘will seemingly, of necessity, be confined to a controversial theoretical standpoint vis a-vis the moral foundations of law’.  

This is particularly important as my goal is to explore a case for regulation that can be reconciled with different accounts of press freedom. While these theories have many differences between them, they are all drawn from a liberal tradition that recognises harm prevention as a legitimate basis for restricting another’s liberty. Many arguments for regulating hate speech are grounded on claims that it causes harm to those it targets. I explore these direct, indirect, and constituted harms as a promising basis for compromise between the libertarian and SRT theorists. Moreover, as Simpson argues, once we establish that hate speech harms, the deontological justifications for restricting it, such as its moral repugnancy, its gratuitous offensiveness, and to condemn reactionary ideas, become more compelling. These various direct, indirect, and constituted harms are explored in later chapters.

III. Balancing

It was noted that the UK is a signatory to ICERD which places on it a positive obligation to proscribe hate speech. The government issued a ‘reservation and interpretive statement’ to the ICERD that explains how it would understand and implement these obligations. It noted that any legislation that it would promulgate would be ‘necessary and proportionate’ and must be balanced against its other human rights obligations in relation to freedom of expression.

In formulating the stirring up offences (which I have noted are narrowly formulated to restrict only overt hate speech), the UK government was attempting to reconcile a

325 Simpson (n 9) 704–705.
327 Brown notes that many of these claims are either explicitly or implicitly predicated on preventing harm Brown, Hate Speech Law (n 55) 50.
328 See chapters 4, 5 and 6.
331 ‘Hate Crime Laws a Consultation Paper’ (n 161) 447.
principle of free speech with the need to secure groups from the various harms caused by such speech. In effect, the government was attempting to arrive at an optimal balance between these conflicting principles. Leonard Sumner captures such balancing in the following terms:

Freedom of expression would be better protected were there no legal constraints whatever on hate propaganda, while the equal status of minority groups would (arguably) be better safeguarded by legislation more restrictive than the hate propaganda law, hedged round as it is by its various safeguards. Somewhere between these extremes lies a balance point at which the greater protection for these groups afforded by more restrictive legislation would be outweighed by the greater impairment of expression, while the greater protection for expression afforded by more permissive legislation would be outweighed by the greater risk of discrimination.332

What Sumner is referring to is a form of balancing that appears prominently in the work of Robert Alexy. To Alexy, principles are not trumps,333 lexical priorities,334 or side-constraints335 but rather ‘optimisation requirements' which should be realised to ‘the greatest extent possible given the factual possibilities.336 In short, when principles collide, they must be balanced with the goal of maximising the realisation of both. He refers to this as the ‘law of balancing’ which requires that ‘the greater the degree of non-satisfaction of, or detriment to, one principle, the greater the importance of satisfying the other’.337 This is the optimal balance that legislators tried to arrive at when framing the stirring up offences.338

This is not to suggest that such balancing is uncontroversial as it has been subject to significant critique. Jürgen Habermas regards it as irrational as it reduces principles (rights) to policy arguments,339 Stavros Tsaryakis argues that conflicting principles are

333 Ronald Dworkin, Taking Rights Seriously (Bloomsbury Academic 2013).
335 Nozick (n 284).
337 ibid 102.
338 For an overview of how Alexy’s theory fits into British constitutionalism see Julian Rivers, ‘Fundamental Rights in the UK Human Rights Act’ in Agustín José Menéndez and Erik Oddvar Eriksen (eds), Arguing Fundamental Rights (Springer Netherlands 2006).
incommensurable as there is no common metric by which to weigh and trade them off, observing that Justice Antonin Scalia\textsuperscript{340} is ‘merely scoring an easy point when he says that we cannot compare the length of a line with the heaviness of a rock’.\textsuperscript{341}

The fear is that balancing will give way to highly subjective judgements. One solution to this subjectivity is proposed by Oster who similarly relies on a balancing approach in fashioning his theory of media freedom. He argues that it can be reduced by setting out ‘rationally reviewable conditions’ that allow for the balancing process to be ‘reasonably predictable’.\textsuperscript{342} The coverage/protection distinction and the category of no, low and high-value provide us with such criteria. Activities that are high-value ought to be assigned great weight in the balancing exercise. By weight here, I simply mean assigned a high degree of importance\textsuperscript{343}, and balancing is simply moral reasoning conducted by legislators in determining the appropriate scope of a statute limiting a principle\textsuperscript{344} and/or by the courts in determining which principle ought to take priority over the other in the circumstances of a particular case.\textsuperscript{345} Classifying speech as high-value creates a presumption in favour of protection or, as Brison puts it, ‘justify weighing the interests with a thumb on the scale in favour of free speech’.\textsuperscript{346} It means that any restrictions on high-value speech ought to be narrowly drawn.

For greater precision, this approach can be supplemented by using the stirring up offences as the reference point for what constitutes the optimal balance for high-value speech. Recall that I am not questioning the justification and scope of hate speech restrictions applying to lone speakers but rather whether the same ones should be applied to the press. Some argue these stirring up offences are too broad,\textsuperscript{347} others that they are too narrow,\textsuperscript{348} and others still that they are just right.\textsuperscript{349} I will assume that proscriptions on overt hate speech achieve an optimal balance between the rights of

\begin{thebibliography}{99}
\bibitem{340} a former judge of the US Supreme Court.
\bibitem{342} Oster (n 24) 127–128.
\bibitem{343} Brown, \textit{Hate Speech Law} (n 55) 228.
\bibitem{344} For a discussion of the merits of a legislative-based balancing see Erik Bleich, \textit{The Freedom to Be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism} (Oxford University Press 2011) 12–13.
\bibitem{345} For a discussion of the merits of judge-based balancing see Oster (n 24) 38.
\bibitem{347} Hare (n 7).
\bibitem{348} Kay Goodall (n 152).
\bibitem{349} Waldron, \textit{The Harm in Hate Speech} (n 14).
\end{thebibliography}
lon speakers and the prevention of harm. They can then be said to reflect the ‘point at which further gains in one of the values will be outweighed by greater losses in the other’. Therefore, either loosening the restrictions to allow a greater degree of hate speech than that which is caught under the stirring up offences or broadening them to catch a broader category of hate speech (such as DNS) would upset this balance.

Low-value speech, on the other hand, is covered and benefits from some protection, but it does not have the same weight as high-value speech and restrictions on it need not be so narrowly drawn. Optimal balancing in such cases requires a presumption in favour of the competing principle of harm prevention by moving up the scale of constraints, i.e. instituting broader bans than the stirring up offences.

We have also seen that no-value speech is not covered by a free speech principle. In such cases, there are no interests to be optimised or balanced, and the state has wide discretion to determine the scope of restrictions to prevent the harm to health, security and dignity occasioned by such speech.

Under this analytical approach, the justifiability of imposing broader restrictions on press hate speech is dependent on whether it should be considered high, low or no-value speech under a free speech principle. Broader restrictions can be justified if press hate speech is either low or no-value speech.

How then do we determine the point of optimal balance when instead, it is a principle of press freedom that is implicated? Oster sets out the criteria for such balancing and draws the line at the same point, arguing that press freedom would not protect the stirring up of hatred. Therefore if hate speech is ‘high-value’ for purposes of press freedom, then it can only be subject to the same narrow restrictions as those imposed on lone speakers. If instead these democratic functions do not require engaging in hate speech, then such speech is of low or no-value under a principle of press freedom, with the same consequences set out above.

---

350 The reasons why this achieves an optimal balance are explored in chapter 7.
351 Sumner (n 332) 62; Alexy (n 336) 102.
352 See chapter 7.
353 Oster (n 24) 233–234.
354 ibid 131.
This is the approach that I will use to answer the central questions. It is a rough guide as balancing and line-drawing are not exact sciences. Waldron points out that it ‘will often be difficult and challenging’.\footnote{Waldron, \textit{The Harm in Hate Speech} (n 14) 115–116.} There will often be a range of acceptable solutions.\footnote{Bleich (n 344) 139.} However, democracy requires that some deference should be extended to representative institutions to determine these finer details.\footnote{Sumner (n 332) 84; Bleich (n 344) 12–13.}

d. The conditions under which broader regulation is justified

From this, I will sketch out the circumstances under which the regulation of DNS would be compatible with the libertarian and SRT accounts of press freedom.

We have seen that the equivalence model does not recognise a distinct principle of press freedom. Rather it assumes the press’s free speech claims are indistinguishable from those of lone hate speakers. It (wrongly)\footnote{See Chapter 7 for a discussion of why this view is unsupported by free speech theory.} assumes that hate speech by the press is high-value speech. Even if we accept this premise, regulation of DNS is still warranted if they produce the same harms at either a greater intensity or scale than those caused by overt hate speech. I sketch out the details of this approach in chapters 4 and 5. I examine the already regulated category of overt hate speech, look at the direct and indirect harms that justify its regulation and determine whether we have empirical evidence of similar quality, demonstrating that DNS by the press causes these harms to a greater degree. In chapter 6, I consider theories of constituted harm built on J.L. Austin’s theory of speech acts and set out why, unlike most lone speakers, the media has the authority necessary for their speech acts to constitute harm. I will argue that constituted harms are particularly problematic as they enact harmful social norms which legitimise and incentivise discrimination.

The variable geometry model, on the other hand, recognises a free speech \textit{and} a press freedom principle, and it must be shown that broader restrictions are compatible with both. We have seen that it subjects press claims for free speech and press freedom to scrutiny under the rationales underpinning the respective principles. Under this model, broader restrictions can be imposed if these rationales suggest:
1. The press ought to have more limited rights to engage in hate speech than lone hate speakers; and
2. These restrictions do not hinder the press’s ability to perform their democratic functions

I will consider these in chapters 7 and 8, respectively. It will be argued that the strongest rationales for treating hate speech as high-value speech are non-instrumentalist speaker-based arguments that either do not—or only marginally apply to—the press. As such, press hate speech is either at the margins (low-value) of, or is entirely uncovered (no-value) by a free speech principle.

In chapter 8, I examine the various democratic functions assigned to the press. I will suggest that, in most cases, these do not require the press to publish hate speech. I will note that there are limited circumstances where these functions do require publishing hate speech, and restrictions can be designed in ways that cater for this.

It will therefore be argued that proscribing DNS is compatible with both the libertarian and SRT theories. The exact nature and scope of such regulation is considered in chapter 9.

4. Conclusion

This chapter sets out the central questions to be examined in this thesis, the methodology relied on and how broader regulation of press hate speech can be justified under each of the prominent theories of press freedom.

Before scrutinising the libertarian and SRT views, I will first address a related issue, which is that regulating DNS in the press would lead to a closer alignment between press regulation and hate speech regulation in broadcasting. This requires justification as a common argument advanced in the media law literature is that the latter is a special medium that ought to be treated differently from the press. I will consider these rationales in the next chapter and whether they provide sound reasons for maintaining the differential treatment of hate speech between the two mediums.
Chapter 3: The Special Status of Broadcasting

1. Introduction

Broadcasters are subject to more far-reaching content regulation of their programming that sets out what can't be included, what must be included, and even how it is included. In chapter 1, we saw that this includes a broader hate speech clause that restricts DNS.

It is claimed that there is something special about broadcasting as a medium which justifies this broader regulation. If this is the case, applying similar hate speech regulation to the print medium may be unjustified if the latter does not also have this special quality to it.

This chapter explores and evaluates various arguments put forth to justify the special status of broadcasting. It will be argued that none provide compelling reasons for the differential treatment of hate speech between these two mediums. As such, the rest of this thesis will proceed on the basis that broader regulation of press hate speech, including DNS, should instead be informed by whether such regulation is compatible with principles of free speech and press freedom.

2. Is the special status of broadcasting justified?

The first question is whether these differences are informed by principles of free speech and press freedom such that a broadcaster's claims under either are somewhat distinct and benefit from different coverage/protection than those by the press. Where, for example, does broadcasting fit within Siebert’s taxonomy? He makes it clear that he intended his theories to apply to ‘all media of mass

359 See for example restrictions on material which glamorises violence Ofcom Broadcasting Code 2020, s 2.4. There is also a 'watershed' that restricts the time of day in which adult programming can be aired Ofcom Broadcasting Code 2020, s 1.4.
360 There are positive obligations placed on Public Service Broadcasters to show 'current affairs and other serious programmes' Communications Act 2003, s 265C; BBC Charter 2016, pt 6 (1).
361 Broadcasters are required to present news and political issues with 'due impartiality' Communications Act 2003, s. 320; Ofcom Broadcasting Code 2020, s.5; BBC Charter 2016, pt 6(1).
362 Ofcom Broadcasting Code 2020, s 3.1 & 3.2
363 Oster (n 24) 137.
364 Wragg, A Free and Regulated Press (n 18) 284.
communication’. Therefore, Siebert includes in his analysis both broadcasting and film (which he explicitly mentions) and arguably current forms of mass communication such as the online offshoots of the press.

It is not suggested that either principle is watered down or applies differently because speech is broadcast rather than printed. Different media can certainly serve different democratic functions, and different democratic models will emphasise some of these roles more than others. However, there is no rationale in assigning different roles and functions purely based on whether the content is delivered through sounds and moving images rather than words.

Rather, the differential treatment of broadcasting is often justified along certain practical and pragmatic grounds as well as its alleged intrusiveness and greater impact. Here I consider whether these rationales provide sound reasons for the broader regulation of hate speech applied in this sector, starting first with the practical and pragmatic grounds (sections a and b) and thereafter its intrusiveness and greater impact (sections c and d).

a. The scarcity of frequencies rationale

The first of these grounds, and which has traditionally had great influence, is the scarcity of broadcasting frequencies. Radio started being used commercially at the beginning of the 20th century, with many stations sprouting up. Siebert notes how this led to chaos as stations, both amateurs and professionals, were broadcasting—and hence overlapping—their signals over the same wavelength bringing this ‘cacophony’ into the audiences’ homes. States needed to step in to solve this.

They intervened by dividing and allocating frequencies in the electromagnetic spectrum for specific purposes, such as the police, military, and other public services, as well as broadcasting. The spectrum available for broadcasting became a scarce

---

365 Siebert and others (n 227) 1.
366 ibid.
367 I consider different models of democracy and their implications on hate speech regulation in chapter 8.
368 Siebert and others (n 227) 84.
resource that could not be allocated to everyone. Broadcasters who were allocated this finite resource effectively became 'public trustees'. This justified both structural regulation in the form of a licensing system as well as content regulation to hold them to these duties. It was in this vein that broadcasters were obligated to provide 'a balanced range of programmes' and a diversity of views for the audience. This is contrasted with press outlets who are not constrained by the same technological limitations. My capacity to publish a paper is not—by dint of technology—diminished by the various other newspapers in circulation. It would therefore be unwarranted to treat the press as a public trustee and impose similar obligations on them.

There are various problems with the scarcity rationale. Firstly, its central premise is that because broadcasting frequencies are in a few hands, it is justifiable to regulate them to ensure the 'widespread dissemination of various points of view'. This means that it is effectively an argument against concentration. However, as Lee Bollinger notes, the problem of concentration is also prevalent in the press, caused not by technological scarcity but by economic pressures and the resultant 'natural monopolization'. This view expressed in 1976 is even more applicable today in Britain's increasingly concentrated press sector where three companies, DMG Media, News UK and Reach, control 90% of newspaper circulation. This rationale does not tell us why scarcity caused by technological constraints is so different from scarcity caused by economic factors that it justifies applying radically different legal regimes to different mediums.

Barendt advances a further argument noting that scarcity can be caused either by the limited frequencies allocated by the government or the limited number of broadcasting stations. He contends that neither supports greater restrictions on broadcasting. Regarding the former, he argues that this scarcity is an 'artificial creation' by the state which has chosen to reserve many frequencies for the police, army, and other public

---

370 Barendt, Freedom of Speech (n 6) 445.
371 Tambini (n 22) ch 3.
374 ibid.
375 Tom Chivers, 'Who Owns the UK Media?' (Media Reform Coalition 2021) 2.
376 Barendt, Broadcasting Law (n 372) 5.
services. It would therefore be illegitimate for them to use the scarcity they created as a justification to restrict the speech of broadcasters. Barendt further contends that the method of allocation also creates scarcity.\(^{377}\) In line with the law of demand and supply in a market economy, the higher the price of a good, the less likely people would demand it. Similarly, if frequencies were allocated to the highest bidder—as they are in spectrum auctions—it is unlikely that there would be an excess of those wishing to broadcast over the supply of frequencies available. Rather, excess demand and scarcity arise where frequencies are awarded at no—or below market—cost.\(^{378}\)

Economic liberals such as Robert Coase, long regarded as the ‘father’ of the spectrum auctions in broadcasting, would take this even further and argue that even scarcity itself does not justify regulation because all resources are finite. He observes that ‘Land, Labor and capital are all scarce…but this of itself, does not call for government regulation’.\(^{379}\) Rather an affirmative case for regulation beyond mere scarcity needs to be made.

Barendt argues that regulation would still be unjustified if scarcity is instead understood as the numerical shortage of broadcasters, as this does not prevent those who are unable to access this medium from conveying their ideas through others.\(^{380}\)

At the time of writing, Bollinger and Barendt saw newspapers as the alternative medium.\(^{381}\) This argument is far more persuasive today due to the accessibility of online video platforms and social media sites which are low-cost avenues to disseminate heterodox views that are underrepresented in the broadcast sector.

Moreover, this scarcity (numerical shortage of broadcasters) was said to have been caused by the then (1961) prohibitively high start-up costs that made the medium inaccessible to most. We would then expect that a scarcity-based regulatory model would focus on the cost of entry rather than the type of medium, but this is not what happens in practice. Barendt observes that it would certainly cost more to set up a

\(^{377}\) ibid 5.

\(^{378}\) ibid 5.


\(^{380}\) Barendt, Broadcasting Law (n 372) 6.

\(^{381}\) Bollinger (n 373) 11; Barendt, Broadcasting Law (n 372) 6.
national newspaper than it would a community radio station, yet it is only the latter which is subject to stricter content standards.\textsuperscript{382}

Bollinger and Barendt's responses to the scarcity rationale are compelling, but they do have some weaknesses. Firstly, even if one were to accept Coase's contention that every resource is finite, it is still the case that some resources are more finite and, therefore, scarcer than others, as Gold is more finite than Iron. Similarly, while the press is constrained by finite resources such as finances and manpower, broadcasting is subject to the same constraints and a limited electromagnetic spectrum. Moreover, while it is arguable that scarcity has been exacerbated by the government allocating only some of the electromagnetic spectrum to broadcasting, it is the case that some trade-off was inevitable here because rival public goods such as national defence and the police require access to the same resource. These are trade-offs that the state is not required to make for newspapers to be published.

Finally, while 'spectrum auctions' have gone from being unthinkable to being a conventional resolution to the scarcity problem, some question the legitimacy of selling what they view as a public resource to the highest bidder.\textsuperscript{383} Others argue that auctions create oligopolies and squeeze out 'free public access and non-profit educational activities'.\textsuperscript{384} It is for this reason that Damian Tambini argues that 'the idea that the market would necessarily be a more “natural” allocation mechanism, or, in the light of externalities and social value, a more efficient one, is simply an ideological construct'.\textsuperscript{385} Once again, even if on balance, the pros of auctions as an allocation method far outweigh the cons, the state is still faced with choices and trade-offs that it is not required to make with respect to the press. While scarcity is not a unique problem to broadcasting, it is still of a greater degree and requires greater trade-offs to resolve. Broadcasting would therefore be special and different in at least this respect.

\textsuperscript{382} Barendt, \textit{Broadcasting Law} (n 372) 6.
\textsuperscript{384} Noam (n 383) 771-772,775.
\textsuperscript{385} Tambini (n 22) ch 3.
Rather, the biggest weakness of the scarcity rationale is that it has been overtaken by technological developments in light of ‘cable, satellite, and digital broadcasting’.\textsuperscript{386} This has led to an increase in the number of ‘actual and potential’ broadcasting stations which now outstrip the number of individual newspapers in circulation.\textsuperscript{387} It is simply the case that the scarcity rationale for the special treatment of broadcasting is no longer sustainable and can certainly not be used to justify broader hate speech regulation.

b. The ‘best of both worlds’ rationale

A pragmatic ground advanced to justify differential treatment is that broadcasting regulation is only ‘tolerable because the press is truly free’. This was a point made by Paul Dacre\textsuperscript{388} in his evidence before the Leveson inquiry, where he noted:

\begin{quote}
Such state involvement [in Broadcasting] only works in a democracy because it is balanced by the commercial press and the internet, both of which are unlicensed and therefore genuinely free.\textsuperscript{389}
\end{quote}

As Wragg points out, the claim being advanced here is that the debate as to whether regulation of broadcasting is ‘compatible with press freedom is not settled but merely suspended for so long as the press and the internet is free from such’ regulation.\textsuperscript{390} The implication is that the broader regulation of broadcasting is probably incompatible with free speech and press freedom, but the press will not raise an issue with the status quo as long as they remain unregulated.

Lee Bollinger is perhaps the biggest proponent of this pragmatic approach, arguing that ‘asymmetric regulation’ of these two mediums is good as it gives us the ‘best of both worlds’.\textsuperscript{391} It allows society to ‘remedy the deficiencies of an unregulated press with a regulated broadcasting system’.\textsuperscript{392} On the one hand, you have regulated

\begin{footnotesize}
\textsuperscript{386} Barendt, Freedom of Speech (n 6) 445.
\textsuperscript{387} Fenwick and Phillipson (n 20) 563; Barendt, Freedom of Speech (n 6) 445.
\textsuperscript{388} The then editor of the Daily Mail and Chairperson of the IPSO editors code committee.
\textsuperscript{390} Wragg, A Free and Regulated Press (n 18) 285.
\textsuperscript{391} ibid.
\textsuperscript{392} Barendt, Broadcasting Law (n 372) 8.
\end{footnotesize}
broadcasters who serve public service functions but are too risk-averse to hold the powerful to account. On the other, you have an unshackled press which is free to 'inform and provoke the agenda in a way the broadcast press cannot'. The weaknesses of one form are compensated for and counter-balanced by the other, creating a mutually reinforcing media ecosystem that better serves the public interest. This rationale, while superficially appealing, is problematic in many respects.

Firstly, Bollinger accepts that broadcasters have the same claims to free speech and press freedom as press outlets. This makes his argument self-defeating and contradictory because if press regulation is wrong (because it is incompatible with free speech and press freedom) and these principles apply equally to both mediums, then surely broadcast regulation would equally be wrong (and equally incompatible with these principles). Similarly, if broadcast regulation is compatible with these principles and is a net good, then there is no reason that the press should not be subject to the same constraints. Cass Sunstein is therefore warranted in noting that although Bollinger's argument to explain the status quo is 'appealing and, even ingenious,' it fails to grapple with 'some hard factual, predictive and theoretical issues'.

Secondly, even if one were to accept that asymmetric regulation leads to desirable ends, it is unclear how this supports the specific kind of asymmetric regulation Bollinger is defending. His argument only makes a case for why some forms of journalism ought to be unregulated to act as a counterbalance to regulated forms. It is not clear how this translates to the current system of regulated broadcasters and an unregulated press and not say the opposite, or why this asymmetry should be based on the medium of communication rather than on a myriad of other characteristics such as circulation numbers, viewership, or revenue. Moreover, it is arguable that even if broadcasting and the institutional press were both regulated, this counterbalance could still be achieved by unregulated citizen journalists and bloggers operating

---

393 Wragg, A Free and Regulated Press (n 18) 285.
394 Cass Sunstein summarises Bollinger’s view as being that looking for differences (between the mediums) is “chasing ghosts” Cass R Sunstein, Democracy and the Problem of Free Speech (Free Press 1995) 122.
395 Barendt, Broadcasting Law (n 372) 8.
396 Sunstein (n 394) 123. Barendt’s formulation is far more succinct, referring to it as a ‘clumsy’ argument Barendt, Broadcasting Law (n 372) 9.
through the internet. The point here is that a rationale that could be used defend the regulation of either medium or none of them is an unappealing basis for the special treatment of a particular one.

Finally, the primary benefit of this asymmetry is that broadcasters are seen as shackled by the threat of ‘governmental control’ and would thus be unable to properly undertake functions such as serious public interest journalism. Lucas Powe Jr makes this point in the context of the Watergate scandal arguing that ‘if the regulated broadcasters were too timid, at least some newspapers were willing to risk potentially serious consequences to stand up to the government’. It is unclear why asymmetrical regulation is necessary to achieve this. Arguably, these goals could also be achieved by a regulatory framework that was both structurally independent and insulated from interference by the government. Ofcom, for example, maintains institutional independence from the state. While a government minister appoints the Chairman and non-executive board members, the executive board members and all other staff are appointed by Ofcom. This could also be supplemented by provisions that provide robust protection for the media when they act as public watchdogs or are engaging in public interest journalism.

The scarcity and best of ‘both worlds’ rationales which are presented as practical and pragmatic justifications for treating broadcasting as a 'special' medium are either entirely overtaken by technological developments or unpersuasive. However, it does not follow that these mediums ought to be subject to the same restrictions. In line with the methodology and analysis set out in chapter 2, the question as to the legitimate scope of hate speech restrictions must be ascertained through balancing. It has been suggested that broadcasting is a medium that is more intrusive and has a greater impact on the audience than newspapers. If hate speech in broadcasting causes greater harm than hate speech in print then, all else being equal, optimal balancing would require imposing greater restrictions on the former, and this would, in theory, support the special status of broadcasting.

397 Sunstein (n 394) 122.
399 Office of Communications Act 2002, sch (Further Provision about OFCOM); Rowbottom (n 8) ch 6.
c. The intrusive medium rationale

The claim that broadcasting has a greater capacity to cause harm than print media has long been used to justify its special treatment. While it is often presented as a single justification, it is in fact comprised of two distinct and severable claims. The first is that broadcasting is a particularly intrusive medium. The second is that there is something unique about how humans process sounds and audio-visual content that makes broadcasting a more powerful and effective medium than words alone. I explore each of these in turn.

It is claimed that television and radio intrude into the home. This is where viewers regularly consume such content, often by tuning in and out of programmes. These consumption patterns mean viewers risk being involuntarily exposed to objectionable content within the privacy of their homes. This view was forcefully set out by the US Supreme Court in *Pacifica* which noted:

> the broadcast media have established a uniquely pervasive presence in the lives of all Americans. Patently offensive, indecent material presented over the airwaves confronts the citizens not only in public, but also in the privacy of their home, where the individual's right to be left alone plainly outweighs the First Amendment rights of an intruder.

This decision has been favourably cited in the UK by Lord Hoffman in *Prolife Alliance*. This intrusiveness is said to be exacerbated by the medium’s pervasiveness and the ease at which children can access it from a terminal at home (and the difficulty in preventing it). Moreover, children can listen and watch at a much younger age than they can read. This is then contrasted with the act of going out to buy a newspaper which is seen as an exercise of choice where the reader voluntarily exposes themselves to offensive content. It is arguable that a person holding liberal attitudes towards immigration who buys a paper that has a xenophobic editorial stance is exercising their autonomy by wilfully exposing themselves to this...
content. However, this autonomy argument is less forceful when the person is exposed to such content while flipping through television channels at home or listening to the radio during their commute.

It becomes even more problematic if such content is encountered by a member of the groups it is directed at. It is one thing for a Muslim to willingly buy and read an Islamophobic paper. It is wholly another for them to be involuntarily exposed to such content while at home with their families. It is this potential to harm privacy and the viewer’s expectations by unexpectedly exposing them to hate speech that is said to justify the broader hate speech restrictions on broadcasters.\(^\text{405}\)

In response to this, Barendt argues that the distinction between what is imposed on an audience and what is chosen by them is difficult to draw as ‘broadcasting does not intrude into the home unless listeners and viewers want it to’.\(^\text{406}\) Moreover, even when the viewer is momentarily subjected to offensive content, they retain the autonomy to change the channel and cease watching the program. Both these arguments are unconvincing. The claim is not that audiences have no choice but rather that they have *less control* over broadcasts accessible from a terminal at home than they do over a newspaper they purchased at a news stand and brought into the home. Secondly, as Justice Stevens noted in *Pacifica* ‘to say that one may avoid further offense by turning off the radio when he hears indecent language is like saying that the remedy for assault is to run away after the first blow’.\(^\text{407}\)

A more compelling response is offered by Phillipson and Fenwick who argue that—like the scarcity rationale—it has been watered down by technological developments. Broadcasting today is far different from what it was when the US and UK courts handed down these decisions in 1978 and 2003, respectively. Viewers no longer simply flip through channels with little information about what to expect. They now have far greater control over what they watch through programme menus that contain descriptions of the programmes as well age ratings, content warnings etc.\(^\text{408}\) The


\(^{408}\) Fenwick and Phillipson (n 20) 564.
medium is now far less intrusive, and consuming such content is arguably as much of an exercise of audience autonomy as buying a newspaper.

Meanwhile, digital technology and the internet have had the opposite effect on the press, which is now far more intrusive. Press outlets now sell both physical papers and publish their content online and through their apps with some, such as the *Independent* successfully shifting to an online-only model. This online content is easily accessible by both adults and children from terminals at their homes (and everywhere else through their portable computers, tablets, and mobile phones). The image of an adult walking to a newsstand and buying a paper is much less relevant today. Moreover, news aggregators that come pre-installed on two of the largest mobile operating systems, Google News (Android) and Apple News (iOS), link to content from press outlets.

Further, all major press outlets in the UK now have an active presence on social media platforms where they promote and link to their material which can be encountered by users who are simply scrolling through their feeds and who had no intention of actively seeking it out. This is exacerbated by the fact that on social media platforms such as *Meta* and *Twitter*, controversial content often becomes a ‘trending topic’, increasing its reach significantly—and therefore increasing the risk of inadvertent exposure. The controversy caused by racist discourse is often amplified by platform features and algorithms, and such controversy can often be generated by the press. For example, the article in *Greer* which compared migrants to 'cockroaches' became a trending topic. It could hardly be argued that users who were exposed to this content and who objected to it were exercising any more choice than they would be when

---

flipping through television channels or radio stations. It is even arguable that the viewer who selects and watches a programme has greater choice and control (through menus, programme details and content warnings) than one who inadvertently encountered this article. The press may now be the more pervasive and intrusive medium. Fenwick and Phillipson were correct in observing that there is a 'degree of tension' between the unregulated internet and the higher level of content standards on broadcasters, more so today.\(^ {414}\) This tension is only likely to be exacerbated if the current government proposals in the Online Safety Bill are actualised. I explore the implications of this proposed legislation in greater detail in chapter 9. Briefly, it seeks to impose a 'duty of care' on certain platforms (this includes social media platforms) to tackle harmful content on their services but explicitly exempts news publishers from its scope for content within their own websites and will include 'robust protection for journalistic content shared' on the social media platform themselves.\(^ {415}\) This may well create a two-tier system where broader hate speech regulation is applied to user-generated content on these platforms and on broadcasters through Ofcom but not on the press.

d. The special impact rationale

It is claimed that broadcasting is more effective in shaping the opinions and attitudes of the audience than print. Lord Hoffman explained this rationale in *Prolife Alliance* where he noted:

*Television in particular makes the viewer feel a participant in the events it depicts and acquainted with the people (real or fictitious) whom he regularly sees. The visual image brings home the reality which lies behind words.*\(^ {416}\)

He stated that this justifies ‘singling out’ television and, to a ‘lesser extent radio’ for special regulation in the form of taste and decency standards.\(^ {417}\) The suggestion is that audio-visual content is the most powerful, followed by audio and then the written word. Similar views have been expressed by the ECtHR on several occasions—albeit

---

\(^{414}\) Fenwick and Phillipson (n 20) 565.


\(^{417}\) Ibid.
they do not draw clear distinctions between television and radio. This was evident in Centro Europa 7 where it was held that:

*The audio-visual media, such as radio and television, have a particularly important role in this respect. Because of their power to convey messages through sound and images, such media have a more immediate and powerful effect than print.*

Perhaps of greatest relevance in the context of hate speech is the court’s decision in *Jersild*, where it was noted that in considering the leeway to be granted to a journalist to disseminate hate speech, the ‘potential impact’ of the medium was an important factor and that the ‘audio-visual medium have means of conveying through images meanings which the print media are not able to impart’. By bringing up the special impact of broadcasting in this context, the court seems to be suggesting hate speech promulgated through this medium has a greater effect on the audience, and journalists must therefore be more careful in what they broadcast than in what they print.

This special impact is said to be made worse by the prevalence of broadcasting as a medium of mass communication. This remains the case in the UK with television still ranked as the most used platform for news, and while radio lags behind, it is still used as a news source by nearly half of adults.

In response to this, Barendt argues that imposing greater restrictions on broadcasting because it is a more effective ‘mode of speech’ is inconsistent with a free speech principle. Elsewhere he notes that ‘it cannot be right to subject more persuasive types of speech to greater restraints than those imposed on less effective varieties’. Barendt here seems to refer to rational persuasion which is based on the assumption

---

419 This is despite the court finding that the restrictions violated Article 10 see Jersild v Denmark [1994] Application number 15890/89, para 31.
420 Bollinger (n 373) 14; Fenwick and Phillipson (n 20) 563.
422 Barendt, *Freedom of Speech* (n 6) 446.
that the person being persuaded is able to rationally and critically reflect on their choices.424

However, commentators and courts referring to a ‘special impact’ seem to be claiming something else. Bollinger points out that what is being suggested is that audio-visual content has ‘some undefined and unquantifiable but nevertheless unique, capacity to shape the opinion of the viewers in ways unrelated to the merits of the argument itself’.425 L.A Powe Jr makes a similar point, noting that the claim being made is that ‘television interacts with all facets of society and changes us in ways that we can neither anticipate nor wholly understand’.426 This view has also influenced more recent work, with Oster arguing that ‘press freedom contradicts any notion of regulation that is content-based and press-specific’ whereas the special impact of broadcasting permits the enforcement of enhanced duties on them, including stricter content regulation.427

This alleged special impact has been influential in various reviews of the broadcast sector in the UK by the Beveridge Committee in 1951, the Pilkington Committee in 1962 and the Annan Committee in 1977, which all treated the unique capacity of television to shape the views and attitudes of the audience,428 to arouse strong emotional reactions,429 and the difficulty faced by the audience in critically analysing ‘transient pictorial statements’ as axiomatic.430 This is perhaps best reflected in this now infamous passage from the Pilkington committee’s report:

*We were told that this effect [on the audience], good or bad, need not be sudden or spectacular. Rather it was to be compared with that of water dripping on a stone: persistent, apparently imperceptible but in the end prevailing.*431

The claim is that there is something unique about the audio-visual medium that gives it the capacity to affect cognition, emotion, and attitudes through non-rational means.

---

425 Bollinger (n 373) 14.
427 Oster (n 24) 137.
430 ‘Report of the Committee on Broadcasting’ (n 428) 15.
431 ibid.
If this is the case, then its special treatment can be justified under a free speech principle. As Barendt himself points out, an assumption of the major arguments for free speech is that people can ‘assess arguments rationally’. If the power of broadcasting is that it can circumvent or lessen the audience’s ability to rationally deliberate on the content being delivered, the same arguments underpinning free speech can justify imposing standards requiring content to be delivered in ways which maximise the audience’s ability to rationally process it. An obvious example would be—as is the case currently—to require some broadcasters to discuss social and political issues with due impartiality rather than a raw appeal to emotion and partisanship. This would enable the audience to carefully assess the respective sides of a contentious issue, thereby minimising the risk posed by the power of the medium. It may also justify broader hate speech regulation in this sector if, for example, viewers are more likely to uncritically accept the various DNS about minorities published through this medium.

A more compelling response to this rationale than that offered by Barendt is simply that it is an empirical claim which is without firm grounding. There are certainly studies linking audio-visual content with a stronger priming effect when it comes to threat perception. For example, moving images are more vivid and stimulating and have been found to elicit more powerful emotional responses than print when covering conflict and violence. However, there is also evidence pointing the other way suggesting print has stronger priming effects. The view is that while audio-visual content is more effective at capturing attention it is more transient, whereas reading takes greater concentration, and the messages in the text are more deeply embedded.

More importantly, modality is just one media property. Media effects scholarship has long examined how the different media properties influence its effectiveness. Patti Valkenburg and others divide these properties into three categories; modality, content,

432 Barendt, Freedom of Speech (n 6) 34.
434 Priming is discussed in depth in chapter 5
and structural.\textsuperscript{436} The claim that audio-visual content has a special impact focuses on modality to the exclusion of these other properties. This is problematic because Valkenburg and others point out that many content and structural properties have a greater impact than modality on how the audience processes information.\textsuperscript{437} Content properties such as the outlet’s perceived credibility,\textsuperscript{438} engaging narratives,\textsuperscript{439} and negative stories\textsuperscript{440}, as well as structural properties such as information cues\textsuperscript{441} and repetition,\textsuperscript{442} all influence the effectiveness of the media. Therefore, a regulatory regime that is geared towards addressing a special impact would consider all these media properties rather than a narrow focus on a single one that is often less consequential than others.

Finally, even if audio-visual content does have a greater impact—sufficient enough to justify its radically different legal treatment—it is no longer unique to the broadcast media.\textsuperscript{443} The press’s online content in their websites and apps integrates audio-visual material into news articles, and yet they are not subject to broadcast regulation. Ofcom has had to develop rules to determine whether such content is within its regulatory ambit (‘in scope’) and, therefore, whether the press ought to be subject to similar obligations as those imposed on broadcasters.\textsuperscript{444} In effect, content is not in scope unless it is presented as an entirely different service, e.g. if a press outlet runs a separate video-on-demand service or if it is completely detached from the article it is

\footnotesize

\textsuperscript{437} ibid 322.


\textsuperscript{440} Valkenburg, Peter and Walther (n 436) 323.


\textsuperscript{443} Barendt, \textit{Freedom of Speech} (n 6) 446.

\textsuperscript{444} For an overview of the various decisions see Francisco Javier Cabrera Blázquez and others, ‘On-Demand Services and the Material Scope of the AVMSD’ (European Audiovisual Observatory 2016) IRIS Plus 2016-1 55–56; Jenny Weinand, \textit{Implementing the EU Audiovisual Media Services Directive: Selected Issues in the Regulation of AVMS by National Media Authorities of France, Germany and the UK} (Nomos Verlag 2018) 323–331.
embedded in.\textsuperscript{445} This means the press have quite a wide latitude to use audio-visual content without complying with the broader content standards imposed on broadcasters, making the special impact rationale ring quite hollow.

3. Conclusion

The scarcity rationale, ‘best of both worlds’ rationale, intrusiveness argument and the special impact of broadcasting either never held up or are no longer relevant. Powe argues that the continued search for an ‘intellectual rationale’ for these differences is really a ‘post-hoc attempt to support the status quo’ (the different regimes between broadcasting and the press).\textsuperscript{446} Rather, these differences are best understood as a matter of historical determinism.\textsuperscript{447} Broadcasting is a (relatively) new medium compared to the press, and just as the cinema is more carefully regulated than theatre, it is understandable that society wanted to carefully regulate broadcasting and to continue to do so even after the presumptive reasons for treating the mediums differently ceased to exist or are inherently weak.\textsuperscript{448} Broadcasters are expected to be impartial and avoid being offensive because this has always been the case. Similarly, the press has not been subject to licensing since 1695, and they are expected to self-regulate, be partisan and free of content regulation. However, as Rowbottom notes, these are good explanations for why we have ended up with two separate legal regimes and not a principled justification for their continued separation.\textsuperscript{449}

As the analysis above demonstrates, no such principled justification exists. While this can be used to justify a wholesale rethink of these different restrictions and to which services they are applied, that is not a claim this thesis will be examining. Many have explored these questions and have considered whether to bring both the press and broadcasters under the ambit of a single regulator.\textsuperscript{450} Such regulators already exist in countries such as Denmark and Finland, where the press council regulates both print

\textsuperscript{445} Blázquez and others (n 444) 56. See also the Sun Video Case <https://stakeholders.ofcom.org.uk/binaries/enforcement/vod-services/sunvideo.pdf> accessed 19 August 2020
\textsuperscript{446} Powe (n 426) 62.
\textsuperscript{447} Barendt, Broadcasting Law (n 372) 9.
\textsuperscript{448} Ibid 8.
\textsuperscript{449} Rowbottom, Media Law (n 42) ch 6.
outlets and broadcasters.\textsuperscript{451} Similarly, Leveson recommended that the current broadcast regulator Ofcom act as a ‘backstop’ statutory regulator of the press if they refused to accept his proposed self-regulatory framework.\textsuperscript{452}

I do not engage with these arguments or proposals in this thesis. Bringing both mediums within the ambit of the same regulator requires examining and determining which of the various duties and responsibilities currently applicable only to broadcasters ought to be applied to the press. Each of these duties and responsibilities requires different sorts of justifications. Arguments for and against impartiality will be different from those for or against pluralism or for restricting the portrayal of violence, and these will differ from arguments for hate speech restrictions. An exploration of the justifiability of applying content standards other than those related to hate speech is both irrelevant to the research questions and impractical to cover within a single thesis.

There could, of course, be a single regulator which sets and applies different content standards for the various media within their purview. This already happens in practice, with Ofcom imposing less stringent content restrictions on in-scope video-on-demand services than it does on public service broadcasters.\textsuperscript{453} However, even this side-steps the issue as the decision not to apply a specific content standard to a particular medium implies that doing so is unjustified. This would also require engagement with the rationales underpinning each of these duties and responsibilities and why it is inappropriate to apply them to some parts of the media.

Lastly, for the reasons touched on earlier (chapter 1), and which I revisit when proposing a specific form of regulation (chapter 9), any attempt to subject the press to mandatory statutory regulation, let alone the same regulator as the broadcast sector would be an exercise in futility.

I focus instead on the normative justifications for applying a specific content standard to the press; regulating DNS. The proposals set out in this thesis will therefore apply


\textsuperscript{452} Leveson (n 2) 1794.

only to the press. The analysis in this chapter has demonstrated that there are no principled, practical, or pragmatic justifications for the differential treatment of hate speech between these two mediums and subsequent chapters will proceed on this assumption. As such, the only question that needs to be answered to justify imposing similar hate speech regulation on the press is whether such regulation can be reconciled with the libertarian and SRT theories of press freedom. A question I turn to now.
Chapter 4: The Direct Harms of Dangerous Negative Stereotypes

1. Introduction

The biggest obstacle to broader hate speech regulation in the press is the libertarian theory preferred by the industry and the equivalence model that it entails. The next three chapters serve two purposes. The first is to set out the various harms that are caused by DNS and which justify its regulation and the second is to demonstrate that such regulation is compatible with the equivalence model.

Section 2 explores the implications of this model, noting that broader restrictions are compatible with it if it can be shown that DNS by the press causes greater harm than overt hate speech by a lone hate speaker. Section 3 explains the different ways that the relationship between hate speech and harm can be conceptualised, and section 4 explores the meaning of causation and sets out the kind of empirical evidence required to demonstrate a connection between speech and harm. Section 5 examines the various direct harms caused by overt hate speech, arguing that some can be caused to a greater degree through DNS by the press.

2. Optimal balancing under the equivalence model

This model assumes that the press’s speech rights are no different to those of the public. In chapter 7, I will argue that there are good reasons for treating hate speech by lone hate speakers as high-value speech. However, even if we accept the equivalence model is correct and that press hate speech is also high-value speech, it does not follow that they should be subject to identical legal treatment. Rather, this only answers half the question; it only tells us that the press’s free speech claims have the same weight as everyone else’s, but not what the optimal balance between this and competing principles ought to be.

My approach focuses on harm prevention, and therefore, the competing principle here is securing the targets of such speech from the various harms it causes. Recall that principles are optimisation requirements and that ‘the greater the degree of non-
satisfaction of, or detriment to, one principle, the greater the importance of satisfying the other’.\textsuperscript{454}

I assume that the stirring up offences which only regulate overt hate speech strike an optimal balance between a hate speaker’s rights and the prevention of harm to their targets. Therefore, if press hate speech is also high-value speech, the only justification for imposing greater restrictions on the press without upsetting this optimal balance is if DNS published by them causes greater harm than the overt hate speech of lone hate speakers. The effect of this would be to increase the weight (relative importance) of satisfying the principle of harm prevention in the balancing process.

Hate speech legislation is enacted with a specific type of speaker in mind, a fringe extremist. Waldron’s ‘loser’ or Dworkin’s ‘eccentrics and cranks’.\textsuperscript{455} The legislative background and history of the stirring up offences, as set out by Dickey, Bindman and Lester, certainly suggest this is the case.\textsuperscript{456} In striking a balance between fundamental rights and harm prevention, legislators had in mind the lone hate speakers’ communicative power and their capacity to cause harm. It will be argued that the press has greater communicative power, and their hate speech can cause greater harm than the hate speech of a lone hate speaker. Optimal balancing in such cases requires scaling up the constraints to catch the speech (DNS by the press) causing these greater harms.\textsuperscript{457}

This requires evidence demonstrating both that DNS published by the press causes the same harms as overt hate speech and that they cause them to a greater degree (scale, intensity and/or likelihood of occurring). To determine this, I will examine the already regulated category of overt hate speech, set out the various harms in the literature that are said to justify its regulation and argue that we have evidence (or at least evidence of similar quality) linking the publication of DNS by the press to these harms. Each of the harms I explore is a sufficient condition for regulating speech. Therefore, there need only be evidence that DNS by the press causes a particular category of harm to a greater degree than the overt hate speech by lone hate

\textsuperscript{454} Alexy (n 336) 102.
\textsuperscript{456} Dickey (n 116); Lester and Bindman (n 116).
\textsuperscript{457} Brown, Hate Speech Law (n 55) 223–224.
speakers. This is not to suggest that I am proposing an incremental extension of criminal law to include proscriptions of press hate speech. Rather, my claim is simply that regulating this content is justified under this balancing approach. For reasons I set out later, criminal restrictions are difficult to reconcile with press freedom.\textsuperscript{458}

There may be other reasons that weigh against regulation, such as ineffectiveness, undesired consequences, a disproportionate cost of enforcement, chilling effect etc.\textsuperscript{459} However, for the moment, assume that it is possible to regulate such content while avoiding these problems.\textsuperscript{460}

3. Hate speech and harm

Many scholars have explored the various harms that are caused by hate speech. Broadly they either posit that hate speech causes harm or it causes and constitutes harm.\textsuperscript{461} It is important to distinguish between these causal mechanisms as it informs the type of empirical evidence one must present to establish the harm in question.\textsuperscript{462} In terms of causal theories, a useful starting point is the recent review of hate speech legislation conducted by the Law Commission which found that theories of harm in the academic literature were centred around the following themes: psychological harm, violence, social exclusion, and marginalisation.\textsuperscript{463}

I will explore these causal harms (and others) over the next two chapters, categorising them into direct and indirect harms. This chapter deals with direct harms, which are those suffered by the victims from reading or hearing the hate speech. I focus on psychological harm and severe emotional distress, as well as misrecognition and silencing. The next chapter considers indirect harms. Hate speech can harm indirectly when a person (A) communicates a message which is heard or read by another person (B) who then engages in conduct that harms a victim (C). For example, where A's speech leads B to have hostility towards, shun, discriminate, or attack C or other members of their perceived ethnic, racial, or religious group. I look at the indirect harms

\textsuperscript{458}See chapter 8
\textsuperscript{459}Simpson (n 9) 705.
\textsuperscript{460}I develop a specific proposal which meets these requirements in chapter 9
\textsuperscript{461}Ishani Maitra and Mary Kate McGowan (eds), \textit{Speech and Harm: Controversies Over Free Speech}\ (Oxford University Press 2012) 6.
\textsuperscript{462}ibid 8.
\textsuperscript{463}‘Hate Crime Laws a Consultation Paper’ (n 161) 69.
of contributing to a climate of hatred and undermining the assurance of security and dignity. When examining each of these harms, I will first look at how overt hate speech causes them, the empirical evidence supporting these claims and whether there is evidence of similar quality to demonstrate that these harms are also caused to a greater degree through DNS by the press.

This empirical evidence comes from a variety of disciplines in the social sciences, including social and cognitive psychology, behavioural economics, communication studies and media effects. These are wide fields of study, and there are limitations in my ability to judge their methodological limitations. I am therefore guided by and draw primarily on Alexander Brown’s examination of the empirical evidence in favour of hate speech laws. I equally rely on Mary Matsuda and Richard Delgado’s work on psychological harm and silencing, as well as Alexander Tsesis’s examination of the link between hate speech and violence. I also draw on original empirical work in this area done by legal scholars such as Katharine Gelber and Luke McNamara. However, I do go beyond this work and supplement it with more recent empirical evidence.

4. Causing harm and empirical evidence: deterministic and probabilistic causation

We have seen that a free speech principle requires speech to benefit from heightened protection. To restrict speech, the harm caused must be sufficiently serious, and the evidence to demonstrate the connection between cause (hate speech) and effect (harm) must be convincing. Despite the importance of demonstrating this connection, and despite the abundance of the causal claims made on either side of the hate speech debate, scholars often fail to elaborate on what they mean by ‘cause’. Broadly, two causal models can be inferred from their work; deterministic or probabilistic causation.

464 Simpson (n 9) 705.
465 Eaton makes this observation in the context of debates about the connection between pornography and harm, but similar claims can be made about hate speech. Further, those who argue that pornography ought to be regulated treat it as either analogous to or amounting to hate speech. AW Eaton, ‘A Sensible Antiporn Feminism’ (2007) 117 Ethics 674, 693–694; Catharine A MacKinnon, Only Words (Harvard University Press 1996).
466 I borrow these terms from Eaton. Those in favour of either model do not identify them in these terms. However, the evidence they suggest is required to establish a causal relationship between speech and harm broadly fits into either Eaton (n 465) 695.
The deterministic model requires evidence that hate speech (\(A\)) is a necessary and sufficient condition for the harm (\(B\)) to materialise, i.e. when we say that \(A\) causes \(B\), we mean that whenever \(A\) occurs, \(B\) occurs. This ‘straightforward material causation’ is often preferred by oppositionists such as Post and Heinze who criticise hate speech laws on the basis that such evidence is lacking.\(^{467}\) Heinze argues that if the harm of hate speech is social intolerance, one would require evidence to show that societies that have more hate speech have greater levels of social intolerance and the opposite trends would be expected in societies with less hate speech.\(^{468}\) He also argues that hate speech bans in Western Europe have not led to a decrease in the levels of racism.\(^{469}\) Similarly, Jacob Mchangama draws on survey data showing that despite the absence of content-based bans, there has been a general reduction in the levels of hate crime and antisemitism in America over the years. He compares this with European states noting opposite trends in these areas.\(^{470}\)

Such comparisons between America and Europe are made for good reason as proponents of deterministic models, including European oppositionists such as Heinze, ‘overwhelmingly endorse’ the American ‘civil libertarian approach to free speech’.\(^{471}\) Such an approach is reflected in the US Supreme Court’s \textit{Brandenburg test} which proscribes speech (including hate speech) that is likely to produce ‘imminent lawless action’.\(^{472}\) Post describes it as requiring a ‘very tight causal connection’ between speech and demonstrable empirical harms.\(^{473}\) John Stuart Mill offers a classic example of this deterministic causal model when he notes:

\textit{An opinion that corn dealers are starvers of the poor, or that private property is robbery, ought to be unmolested when simply circulated through the press, but may justly incur punishment when delivered orally to an excited mob assembled before the house of a corn-dealer.}\(^{474}\)

\(^{467}\) Heinze, \textit{Hate Speech and Democratic Citizenship} (n 5) 153; Post, ‘Hate Speech’ (n 15).


\(^{469}\) Heinze, \textit{Hate Speech and Democratic Citizenship} (n 5) 153.


\(^{471}\) Post, ‘Hate Speech’ (n 15) 194.

\(^{472}\) Brandenburg v. Ohio, 395 US 444 (1969)

\(^{473}\) Post, ‘Hate Speech’ (n 15) 134.

\(^{474}\) Mill (n 287) 94.
This model also appeals to proponents of at least some limited forms of viewpoint punitive hate speech laws such as Barendt. He argues that at least where political speech is concerned, the most appropriate approach is to require the harms to be ‘certain and unavoidable’.\(^\text{475}\) Put simply, the evidence must show \textit{but for} hate speech, the harm would not have occurred.\(^\text{476}\) Waldron is critical of this model, arguing that oppositionists ‘raise the bar very high’, requiring us to ignore evidence ‘unless it is established beyond a scintilla of doubt’.\(^\text{477}\) I agree and consider the deterministic causal model to be unsuitable for two reasons.

Firstly, it is unsuited to explain human behaviour. Deborah Cameron and Elizabeth Frazer liken the deterministic model to billiard balls whose behaviour is ‘determined by the laws of physics’. One simply needs to know the properties of the balls and that of the environment to accurately predict their trajectory, angle, and speed when they are struck, and billiard balls behave this way in every context.\(^\text{478}\) This model also requires that the harm of hate speech reliably manifests in the same way in every context. Cameron and Frazer point out that while the deterministic model may be useful in the realm of physical sciences, it is inappropriate to explain human behaviour as ‘humans are not like billiard balls’.\(^\text{479}\) Unlike the balls, human response to stimuli (in this context, hate speech) is not linear. We interpret, interact and engage with the speech to ‘produce meaning from it’.\(^\text{480}\) This meaning can vary significantly based on a plethora of factors such as personality traits, temperament, mood, values etc. which the speech is filtered through. Some people who are directly targeted by hate speech might respond as Baker suggests they should by ignoring the hate speaker and ‘maintaining and affirming their identity’ in the face of racial abuse.\(^\text{481}\) Others may suffer psychological harm, several may respond with anger, and some may feel nothing at all. The deterministic model is even less appropriate for indirect harm, where we must account for the effect of hate speech on the listener and account for the listener’s

\(^{475}\) Barendt, \textit{Freedom of Speech} (n 6) 170.

\(^{476}\) Frederick Schauer, ‘The Phenomenology of Speech and Harm’ (1993) 103 Ethics 635, 642.

\(^{477}\) Waldron, \textit{The Harm in Hate Speech} (n 14) 148.


\(^{479}\) Itzin (n 478) 249.

\(^{480}\) ibid 251.

\(^{481}\) Baker, \textit{Human Liberty and Freedom of Speech} (n 237) 992.
subsequent attitudes and behaviour towards the targeted group, introducing a plethora of intervening variables between cause and effect.

This means that speech is rarely—if ever—a *but for* cause of harm.\textsuperscript{482} Take Mill’s corn-dealer example. The causal claim (violence against the corn-dealer) is reliant on various conditions, such as the mob’s prior dislike of the corn-dealer(s), a proclivity for violence, and their indifference to the morality of their actions or their legal consequences.\textsuperscript{483} Goodall is correct in noting that if the deterministic model preferred by some oppositionists were to be taken seriously, even requirements such as fighting words or imminent lawless action would fall short of this standard.\textsuperscript{484}

Secondly, even if the deterministic model is appropriate, it is unsuitable for the purposes of my analysis because it does not correspond to the kinds of hate restrictions found in European states, including the UK. Recall that in this thesis, I am not questioning the appropriateness of the scope of current criminal law restrictions on hate speech as they apply to the public. Those debates have been had and will continue to be had. I am operating under the assumption that they are appropriate and justified. To restrict DNS, there need only be sufficient evidence to satisfy the causal model used to justify current bans on overt hate speech. What then is this causal model?

UK hate speech laws accept a less proximal connection between speech and harm than that required by the deterministic model. Recall that the stirring up offences proscribe speech that is likely to incite hatred. The stirring up of hatred is the offence, no concrete act—let alone an imminent one—needs to result from the speech.\textsuperscript{485} Post compares this kind of legislation to the ‘bad tendencies’ test that was deployed in the US prior to the *Schenk*\textsuperscript{486} and *Brandenburg* where speech was restricted based on its ‘general tendency to cause social harm’ regardless of whether any such harm resulted in a specific case.\textsuperscript{487} Stefan Sottiaux argues that the bad tendency test has also been

\textsuperscript{482} Schauer, ‘The Phenomenology of Speech and Harm’ (n 476) 642.
\textsuperscript{483} I borrow this analogy from Schauer with necessary modifications ibid 643.
\textsuperscript{484} Goodall (n 17) 221.
\textsuperscript{485} ibid 213–214.
\textsuperscript{486} Schenck v. United States, 249 US 47, 52 (1919).
\textsuperscript{487} Post, ‘Hate Speech’ (n 15) 134.
embraced by the ECtHR. Unlike the deterministic model, the bad tendencies test treats causation as indirect and contingent.

This test more closely resembles a probabilistic causal model. Eaton argues that our inability to make ‘total predications about human behaviour and psychology’ means that we must settle for ‘probabilistic approximations in the domain of human conduct’. Probabilistic causation has been conceptualised in different ways. But all conceptions have this basic premise. We can say that $A$ is a cause of $B$ if $A$ is a salient risk factor of $B$. In other words, the presence of $A$ increases the likelihood of $B$, but it does not guarantee that $B$ will occur. $A$ is neither a necessary nor sufficient condition of $B$’s occurrence.

This is also reflected in Brown’s ‘precautionary principle’, which requires ‘some minimally adequate evidence’ that hate speech has certain ‘potentially harmful effects’. He argues once this is established, the burden of proof is shifted to those who wish to permit such speech to prove that it ought not to be regulated ‘based on evidence that is sufficiently rigorous, comprehensive, and abundant to command a consensus among the relevant body of experts’. Therefore, this precautionary principle is triggered—and hate speech can be restricted—once it is demonstrated that hate speech is a contributing factor to various harms. This certainly reflects the legislative background of the stirring up offences as when introducing the Bill which created them, the Home Secretary noted that the stirring up offence was intended to ‘deal with more dangerous, persistent, and insidious forms of propaganda campaigns—the campaign [sic] which, over a period of time engenders hate which begets violence’. The legislation was therefore framed in precautionary terms.

---

489 Goodall (n 17) 220; Brown, Hate Speech Law (n 55) 26.
490 Eaton (n 465) 697.
493 ibid 613.
494 Simpson (n 9) 707.
I will adopt the same approach to causation when evaluating the empirical evidence. To support broader regulation of the press, there must be evidence that DNS are contributing factors to the various harms that have justified restrictions on overt hate speech, and that they do so to a greater degree than the hate speech of a lone hate speaker.

I will therefore rely on those studies which suggest that there is a statistically significant relationship between hate speech and the various direct and indirect harms I explore. Any references to ‘cause’ in the remainder of this thesis will therefore be understood as ‘contributing factor’. This model of causation and the bad tendencies test will not satisfy all oppositionists, but it is the model upon which current bans are based and, therefore, the most appropriate model to inform my approach. I will now use it to evaluate the direct harms caused by overt hate speech and the extent to which these are also caused by the publication of DNS by the press.

5. Direct harms: Targeted and diffuse hate speech

When evaluating direct harms, it is important to distinguish between targeted and diffuse hate speech.497 Targeted hate speech is that which is directed at specific individuals. This could either be in face-to-face encounters, such as shouting a racial epithet at a person498 or by delivering the hate speech through both offline (personal letters, graffiti on their doors) and online (direct messages on social media, emails) mediums.499 It also includes targeting a group of individuals or small communities within a larger social group, such as Barendt’s example of racist leaflets strategically left in places where members of the community congregate.500

At the other end of the spectrum is diffuse hate speech which is directed at entire groups and disseminated to the general public. What Waldron would describe as those messages that are a ‘semi-permanent or permanent form of the visible

---

497 I borrow this from Caleb Yong who distinguishes between targeted and diffuse vilification. I do not use the term vilification as I examine a broader genera of hate speech Caleb Yong, ‘Does Freedom of Speech Include Hate Speech?’ (2011) 17 Res Publica 385.
498 This is often referred to as ‘face-to-face vilification’ Brison (n 346) 313; Heinze, Hate Speech and Democratic Citizenship (n 5) 127.
500 Barendt, Freedom of Speech (n 6) 175.
Examples include defacing public buildings with epithets, public posts on internet forums or social media platforms, and the publication of messages denigrating groups in newspapers or broadcasting them through the radio or television.

a. Psychological harm and severe-emotional distress

It is often claimed that overt hate speech causes its targets to suffer psychological harm and severe-emotional distress. Whether certain psychological and emotional states qualify as harms is disputed, Brown notes that it is unclear whether they would fall within Mill’s harm principle, but excluding them would also be unduly restrictive. 502 Joel Feinberg treats transient forms of discomfort such as anxiety, irritation, discomfort, and boredom as falling outside any meaningful definition of harm. 503 Importantly, Feinberg does not entirely rule out psychological states from the harm principle, noting that mental suffering that is ‘severe, prolonged or constantly repeated’ would qualify as harmful. 504 Similarly, Erik Bleich writes that ‘intense psychological pain’ and emotional distress ought to be legally actionable harms. 505

What then are these psychological states that are severe enough to count as harms? Melina Bell argues that with the benefit of empirical evidence that was not available to Mill, we have been able to link hate speech to ‘forms of tangible psychological harm’, a view echoed by Brown. 506 To support their claims, both refer to the work of the critical race theorists Mari J. Matsuda, Richard Delgado, Charles R. Lawrence III, Jean Stefancic and Kimberlé Williams Crenshaw. 507 These scholars argue that the immediate effect of hate speech is severe emotional distress experienced as ‘flinching, tightening of muscles, adrenaline rushes and an inability to sleep’. 508 These harms are

501 Waldron, The Harm in Hate Speech (n 14) 37.
502 Brown, Hate Speech Law (n 55) 51.
503 Feinberg, The Moral Limits of the Criminal Law (n 326) 47.
504 ibid 45–46.
505 Bleich (n 344) 146.
508 Delgado (n 507) 13–14.
caused by the shock and trauma of being subject to interpersonal abuse and are more likely to manifest in targeted hate speech, particularly in face-to-face contexts where it carries an implicit threat of violence. Lawrence describes these as ‘verbal slaps’, signifying the similarity between targeted overt hate speech such as ‘you goddamn nigger, go back to Africa’ and physical assaults.

It is also argued that these verbal assaults cause short-term and long-term psychological harm, such as depression, repressed anger, and self-hatred. Delgado notes that these mental states can have real-world consequences on the victim’s ability to maintain relationships and employment. Victims may, in turn, resort to antisocial behaviour and/or substance abuse as escape mechanisms.

Heinze and John Bennett question the veracity of the empirical evidence underpinning these claims. Some of these differences are based on their preference for a deterministic causal model. Others are more fundamental. Heinze casts doubt on the rigorousness of the early work by Matsuda and others ostensibly on the grounds that they were first published in ‘in-house university’ journals which at the time were not subject to standard peer-review. Bennett and Bleich, on the other hand, question the age and contemporary relevance of the psychology literature relied on by Delgado. The issue with both these criticisms is that these harms have, over the years, been replicated in various peer-reviewed studies. The short and long-term

---

509 ibid 12.
511 Matsuda (n 11) 91.
512 ibid.
513 ibid.
515 Heinze, Hate Speech and Democratic Citizenship (n 5) 127.
516 Bennett (n 514) 488; Bleich (n 344) 146.
impacts of identity-based acts of discrimination, including hate speech, on a victim's mental health, are now well documented.518

Bennett also questions whether victims who claim to be suffering these harms may either be confused as to the cause of their psychological state, misinterpreting the intention of the speaker, or overstating the harm.519 These are reasonable concerns. These studies are heavily reliant on a victim’s perceptions and memories of past exposure to hate speech, and this certainly raised questions of epistemic validity. As Brown notes, first-person accounts from memory ‘could be susceptible to underreporting, exaggeration or misidentification’.520 However, such evidence may be all that can realistically be obtained. As many psychologists have noted, there are substantial ethical constraints that prevent intentionally exposing people to hate speech and measuring their reactions against that of control groups.521 These researchers were also not blind to these methodological criticisms, which they directly address and mitigate in their studies.522

Therefore, there is sufficient evidence to demonstrate that targeted overt hate speech is at least a contributing factor to psychological harm and severe-emotional distress. Laws that catch this subset of hate speech, such as the racially or religiously aggravated public order offences which criminalise targeted overt hate speech that is intended to523 or likely to524 cause harassment, alarm, or distress, appear to be empirically justifiable.

This also lends support to the proposed Harmful Communications Offence in the proposed Online Safety Bill, which criminalises the sending of a message that creates a ‘real and substantial risk’ of harm to a ‘likely audience’ with harm being defined as ‘psychological harm amounting to at least serious distress’.525

518 Nelson (n 57) 10; Neville Lawrence, The Routledge International Handbook on Hate Crime (Nathan Hall and others eds, 1st edn, Routledge 2014) 41–43.
519 Bennett (n 514) 488–489.
520 Brown, Hate Speech Law (n 55) 55.
523 Crime and Disorder Act 1998, s.31 (1) (b); Public Order Act 1986, s.4A
524 Crime and Disorder Act 1998, s.31 (1) (c); Public Order Act 1986, s.5
525 Online Safety Bill 2022, s. 150 (1), (4).
b. Psychological harm in diffuse press hate speech: the absence of empirical evidence

I have already identified that press hate speech is primarily diffuse (DNS directed against groups), but there are a few cases where the press has engaged in targeted hate speech.

One would expect the individuals subject of such speech to suffer similar short and long-term psychological harm and emotional distress, particularly as, unlike a verbal assault by a lone hate speaker, the victims here are attacked before a larger audience. The complainants in the IPSO decisions in Manji\(^{526}\) and Trans Media Watch\(^{527}\) both described the distress and humiliation they experienced from being individually targeted by Islamophobic and Transphobic reporting.\(^{528}\)

However, we have seen that press codes already prohibit targeted hate speech by restricting pejorative and prejudicial references to an individual, and any failure to deal with this content is a failure of enforcement rather than the absence of standards.\(^{529}\) The inclusion of these provisions protecting individuals from targeted hate speech in both press codes is certainly grounded in empirical evidence. To justify also restricting diffuse press hate speech, i.e. DNS, we would require evidence of similar quality that diffuse hate speech also causes psychological harm and severe emotional distress and does so to a greater degree. Such evidence is not forthcoming.

Firstly, none of the studies explored thus far link diffuse hate speech with severe emotional distress. Delgado explicitly notes that the harms caused by this speech are ‘qualitatively different’ from the face-to-face variety.\(^{530}\) Recall that severe emotional distress is caused by the shock and trauma from being targeted with a verbal assault, and these harms are therefore only likely to materialise in interpersonal contexts.

---

\(^{526}\) Manji v The Sun 05935-16 (IPSO).

\(^{527}\) Trans Media Watch v The Sun [2015] IPSO 00572-15


\(^{529}\) IPSO Editors Code, Clause 12; IMPRESS Standards Code, Clause 4.1

\(^{530}\) Delgado (n 507) 12; Richard Delgado and Jean Stefancic, ‘Four Observations About Hate Speech’ (2009) 44 Wake Forest Law Review 353, 363.
There is equally an absence of evidence that diffuse hate speech causes psychological harm. For example, a study by Brendesha Tynes and others found a causal connection between targeted hate speech in online communication and an increase in levels of depression but failed to find a similar connection with diffuse hate speech.\(^{531}\) Another well-known study by Howard Ehlrich and others examines the psychological effects of exposure to both targeted and diffuse overt speech. It concluded that the highest rate of symptoms of post-traumatic stress was amongst those who were victims of targeted hate speech.\(^{532}\) Those exposed to diffuse hate speech were less likely to experience these symptoms, but they still experienced them at higher rates than the control group (who had not encountered any hate speech).\(^{533}\)

However, Ehrlich’s study is limited to overt hate speech and includes examples such as a spray-painted Ku Klux Klan logo in a public place.\(^{534}\) None of the studies thus far have been able to link covert hate speech with these harms. Part of this is simply due to a paucity of empirical literature on covert hate speech generally, but even studies which do examine it point to victim responses such as feelings of shame, which fall short of psychological harm.\(^{535}\)

Delgado and Stefancic would argue that this is because covert hate speech is generally not as harmful as overt hate speech.\(^{536}\) They claim that this is because overt hate speech is perceived by the victim as a more serious wrong because it is plainly ‘calculated to offend’, whereas they find the more moderate language used in covert hate speech to be less shocking, less frightening, and less disturbing.\(^{537}\)

Matsuda, on the other hand, takes a broader view recognising that in some contexts, covert hate speech can cause psychological harm and severe-emotional distress. She uses the example of holocaust denial, arguing that (whether covert or overt) it reminds Jews of a history of discrimination and violence and is always recognizable by them.

---

532 Perry (n 522) 156.
533 ibid 156–158.
534 ibid 155.
536 Stefancic and Delgado (n 507) ch 5.
537 Delgado (n 507) 38–39.
as an expression of serious contempt.\textsuperscript{538} Matsuda refers to it as defamation of the dead.\textsuperscript{539} It has been referred to in similar terms in ECtHR jurisprudence which treats holocaust denial as the ‘the most serious forms of racial defamation of Jews and of incitement to hatred of them’\textsuperscript{540} that justifies criminalising even its covert forms ‘dressed up as impartial historical research’\textsuperscript{541}

It may also be that because of this history of trauma, the targets of covert antisemitism may be uniquely prone to suffering psychological harm. Brown draws a useful analogy to the ‘egg shell skull’ rule in tort law and argues that ‘hate speakers must take their addressees as they find them’.\textsuperscript{542} There still exists today a significant number of direct victims of the holocaust and a much higher number of second and third-generation survivors, and many Jews have both close and distant family members and friends who were direct victims.\textsuperscript{543} While there is some uncertainty in some areas of the literature on the effect of transgenerational transmission of Holocaust trauma, there is a consensus that second-generation survivors have high rates of Post-traumatic stress disorder (‘PTSD’), and offspring of parents who have PTSD have higher rates of anxiety, difficulties in coping with stress and are more likely to develop PTSD themselves.\textsuperscript{544} To these people, it is conceivable that exposure to this literature, even where it is ‘devoid of explicit hate language’, which denies the very source of their trauma, would cause psychological wounds similar to a racial epithet.\textsuperscript{545} However, outside of these narrow historically contingent circumstances, there is indeed a paucity of evidence linking covert diffuse hate speech with psychological harm.

There is certainly a need for more empirical research to investigate this link (between covert hate speech and psychological harm). However, as matters stand, the paucity of evidence means that one cannot offer a convincing argument that DNS by the press causes similar, let alone greater, psychological harm and severe emotional distress

\textsuperscript{539} Matsuda (n 11) 42.
\textsuperscript{540} Garaudy v. France [1998] (Application no. 65831/01), para 23.
\textsuperscript{541} Perinçek v. Switzerland [2013] (Application no. 27510/08), para 243.
\textsuperscript{542} Brown, Hate Speech Law (n 55) 56.
\textsuperscript{543} See generally William B Helmreich, Against All Odds: Holocaust Survivors and the Successful Lives (Routledge 2017).
\textsuperscript{545} Matsuda (n 11) 41–42.
than that produced by overt targeted hate speech. An argument for regulating DNS cannot be predicated on these harms.

c. Misrecognition, self-esteem, and silencing

Another promising candidate rationale for regulating hate speech is that it amounts to misrecognition, which lowers the self-esteem of the targeted groups and silences them. There are different versions of the ‘silencing argument’, and these are often based on constituted harms which I explore in a later chapter. However, the version relevant to direct harms was first advanced by critical race theorists who argue that targeted overt hate speech silences its targets in two ways. Firstly, there is a short-term silencing. Lawrence argues that the ‘fear, rage and shock’ of interpersonal racist abuse debilitates victims into silence. Because this verbal abuse renders the victim speechless, or at least unable to muster a rational response, it is said to harm their free speech interests and denies any bystanders the opportunity of hearing them.

The claim is not that this is a legal violation of the speaker’s First Amendment rights—as this is generally only a negative right from state interference—but rather that even in the absence of legal restrictions, disadvantaged minority groups routinely have their free speech curtailed by face-to-face racist abuse. Put simply, Lawrence’s claim is that if the underlying justification of the right is open debate, minorities are hindered from participating in it by racist abuse, which in his words, reduces ‘the total amount of speech that reaches the market’. He suggests—perhaps counterintuitively—that more speech may require more regulation or, as Owen Fiss puts it, ‘sometimes we must lower the voices of some in order to hear the voices of others’. This approach to harm is compelling in that the purpose of restrictions is to optimise not between free speech and dignity or security but rather between competing free speech interests. Of course, to fully understand why silencing is harmful, one would have to make an affirmative case for why free speech is valuable to the individual and

546 See chapter 6.
547 Matsuda (n 11) 68.
548 Ibid.
550 Matsuda (n 11) 79.
why its curtailment harms them. These justificatory theories are fleshed out later in this thesis. Suffice it to say that we value free speech for many reasons, and what the silencing argument demonstrates is that these values appear on both sides of this issue.

However, the problem here is the transient nature of short-term silencing. In the cases Lawrence describes (face-to-face vilification), the potential audience members, particularly in offline contexts, are the hate speaker and a few other members of the public who are within earshot of the verbal assault. If silencing has only a transient effect, it would not seem to meaningfully diminish the victim’s ability to participate in wider public debate, and the victim has not therefore been harmed, whereas legal restrictions on hate speech are permanent limits on the hate speaker’s speech.

This is where long-term silencing comes in. It is also claimed that targeted overt hate speech lowers the victim’s self-esteem, causing them to either self-censor or withdraw entirely from public discourse.\(^{552}\) If this is the case, the silencing argument has more teeth as it effectively means that a state’s failure to proscribe overt hate speech allows the creation of an environment where minorities are routinely abused into silence by hate speakers and excluded from participating in public debate altogether.

The impact of this silencing effect can be exacerbated by the population sizes of the group in question. For example, in England and Wales, Black Caribbeans account for roughly 1% of the population and retreating to purely in-group interaction effectively amounts to an exclusion from the public sphere.\(^{553}\) It can also create a vicious cycle where hate speech causes minorities to avoid the outgroup; this increases the outgroup’s hostility towards them which then causes further outgroup avoidance.

How then does targeted overt hate speech cause this silencing? The clearest explanation in this early work is offered by Delgado, who argues that hate speech lowers the victim’s self-esteem by causing them to internalise the ‘traits of inferiority

---

\(^{552}\) Matsuda (n 11) 68–69.

that it ascribes’ to them, e.g. that they are stupid or lazy. These claims have been fleshed out in subsequent literature, which conceptualises this loss of self-esteem as a by-product of misrecognition.

There is a growing body of prohibitionist work that relies on recognition theory to defend hate speech restrictions. Steven Heyman relies on Hegelian ideas of mutual recognition to argue that hate speech violates the victim’s right to recognition. Simon Thompson draws on Axel Honneth’s theory of recognition to defend the stirring up offences. Similarly, Brown utilises the three forms of recognition developed in Charles Taylor’s *the Politics of Recognition* to argue that restrictions on hate speech are justified if they ‘limit misrecognition and promote recognition’.

Misrecognition spells out precisely how hate speech causes a loss of self-esteem. It is argued that our identity, understood here as our perception of who we are, is to some extent, socially derived. Put differently, in order for us to develop a sense of who we are, we must be recognised by others. Honneth notes, ‘one can develop a practical relation-to-self only when one has learned to view oneself, from the normative perspective of one’s partners in interaction, as their social addressee’. It is because of this ‘internal interdependence’ on recognition from others that misrecognition can have devastating consequences on our self-esteem.

Misrecognition occurs when the image of ourselves reflected back to us by others is one that is ‘confining, demeaning or contemptible’. Taylor argues that the internalisation of these images can also ‘inflict a grievous wound, saddling its victims with a crippling self-hatred’. Self-hatred is destructive to the victims’ well-being in a multitude of ways. It affects the quality of their lives, undermines their prospects for

---

554 Matsuda (n 11) 94–95.
557 Brown, *Hate Speech Law* (n 55) 167; Parekh, *Rethinking Multiculturalism* (n 57).
559 Thompson (n 556) 224.
561 ibid 116.
562 ibid 116.
563 ibid 26.
564 Thompson (n 556) 228.
self-realisation, and limits their ability to create and maintain relationships with others. It also informs us precisely how hate speech lowers self-esteem; by causing minorities to internalise the belief that morally irrelevant characteristics such as race and sex are ‘distinctions of merit, status and personhood’ or, as Parekh writes, ‘it demeans them in their own eyes’.66

To preserve their self-esteem, members of minority groups are left with two options. They can either isolate themselves from members of the majority by retreating to and exclusively interacting with their in-group. Alternatively, they can distance themselves from their own in-group and erase their own personhood to blend in with the majority in the hopes of preventing similar attacks in the future. Both options silence, the former by excluding them from public debate, the latter by forcing them to erase their identity and adapt their preferences to fit in with the outgroup.

Notably, there is empirical evidence establishing a causal link between targeted overt hate speech and these reactions from victims. Various studies on the experiences of victims of overt targeted hate speech found that these slurs often led to a loss of self-esteem and outgroup avoidance. Research also demonstrates that it causes victims to overestimate the extent to which discriminatory attitudes are held by the outgroup, leading to what social psychologists refer to as ‘cultural mistrust’. This clouds their interactions with members of outgroups by making them more guarded, compounds their feelings of ostracization and causes them to self-censor. It can therefore be

565 Brown, Hate Speech Law (n 55) 126.
566 Matsuda (n 11) 91.
567 ibid 90; Waldron, The Harm in Hate Speech (n 14) 84.
568 Parekh, Rethinking Multiculturalism (n 57) 314.
569 Bhikhu Parekh, ‘Is There a Case for Banning Hate Speech?’ in Michael Herz and Peter Molnar (eds), The Content and Context of Hate Speech: Rethinking Regulation and Responses (Cambridge University Press 2012) 45.
570 Brown, Hate Speech Law (n 55) 84.
571 Matsuda (n 11) 91; Brown, Hate Speech Law (n 55) 198.
claimed that overt targeted hate speech contributes to a lowering of self-esteem and silencing.

d. The media’s unique capacity to silence: The phenomena of stereotype threat

Here I will argue that DNS by the press also silence and that the communicative power of the press gives them a greater capacity to do so than an ordinary or even influential hate speaker. If this is the case, then restrictions on DNS can be justified under the equivalence model.

Starting with its diffuse nature, there is no reason to believe that targeted hate speech lowers a victim’s self-worth and silences them, whereas diffuse hate speech directed against their social groups does not. Brown aptly refers to the harm of misrecognition as the tarnishing of an identity. Once the identity of a group is tarnished, anyone whose personal identity is based on their social identity, ‘I am a proud Muslim’, will have a diminished sense of self-worth. Further, even those who do not identify with their social group are still affected by the tarnishing of the group’s identity because others identify them with it.

The question is whether these harms also manifest in DNS by the press, which, as I have noted, is a form of covert hate speech. I argue that they do for two reasons. Firstly, recall that long-term silencing is caused by the victim coming to recognise and internalise the messages of inferiority communicated by hate speech. We have seen that one way of communicating these messages is through overt hate speech, such as racial slurs. However, note here that the slur is simply the mode of delivery. Rather, it is the DNS that these groups are ‘stupid, lazy, dirty, or untrustworthy’ whose internalisation leads to the loss of self-esteem. These DNS can also be communicated using moderate language that does not breach civility norms. This is not a novel argument. Feminist legal scholarship, notably in the work of Rebecca

575 Brown, Hate Speech Law (n 55) 169–170.
576 ibid 169.
577 ibid 168; Afshin Ellian and Gelijn Molier, Freedom of Speech under Attack (Eleven International Publishing 2015) 266.
578 ibid 169.
579 Maitra and McGowan (n 461) 236–237.
Cook and Simone Cusack\textsuperscript{580} and more recently by Alexandra Timmer,\textsuperscript{581} has recognised the power of gender stereotyping to diminish women's self-esteem, with both explicitly referring to stereotyping as a form of misrecognition.

Secondly, there is some empirical evidence in the legal literature and even more in media effects literature that supports this claim. In legal scholarship, Gelber and McNamara’s study interviewed groups who were vilified by the media, and they found that for some, this coverage led to feelings of helplessness, exclusion from the rest of the community\textsuperscript{582}, and therefore silencing.\textsuperscript{583} These findings broadly fit in with conclusions drawn in media effects studies which have used the framework of stereotype threat theory to examine the effect of being exposed to stereotypes about one’s own group.\textsuperscript{584} Stereotype threats occur when individuals believe that their ingroup is negatively evaluated by the outgroup.\textsuperscript{585} Muniba Saleem and Srividya Ramasubramanian note that the prevalence of negative cultural stereotypes promulgated by the media is one of the ways that the in-group is devalued.\textsuperscript{586} Several studies on stereotype threat in the media have found that such hate speech causes silencing. Saleem and Ramasubramanian found that Muslim Americans exposed to media containing Islamophobic stereotypes have an increased desire to avoid the outgroup, irrespective of whether the participants had personal experiences of discrimination.\textsuperscript{587} Another study suggests that exposure to media containing negative stereotypes about one’s group caused a loss of self-esteem.\textsuperscript{588}

\textsuperscript{580} Rebecca J Cook and Simone Cusack, \textit{Gender Stereotyping: Transnational Legal Perspectives} (University of Pennsylvania Press 2011) 59.


\textsuperscript{582} Gelber and McNamara (n 535) 334.

\textsuperscript{583} ibid.


\textsuperscript{587} ibid.

A study on Dutch media discourse interviewed various immigrant communities who were subject to DNS by the media, which labelled them as criminals and terrorists.\textsuperscript{589} The study noted that some of the most common reactions from members of the targeted communities were disillusionment with Dutch life or distancing themselves from their own group.\textsuperscript{590} Dissociation with one’s identity was also noted in a study of Somali women in London, attributable primarily to what the researchers described as ‘racist anti-immigrant discourses in the media’.\textsuperscript{591} These responses are remarkably familiar because they are the same ones observed by Matsuda and others in the context of overt hate speech. The similarities do not end there as stereotype threat literature also theorises the victim groups’ responses as coping mechanisms to maintain self-esteem.\textsuperscript{592} Finally, it creates greater cultural mistrust than the overt hate speech of a lone hate speaker. Minorities are more likely to believe that the views expressed in a newspaper are more widespread than the beliefs of lone extremists. It is more likely to ostracise them and cloud their interactions with members of the outgroup by making them more guarded and therefore self-censor to a greater degree.\textsuperscript{593} The ability of covert hate speech to silence is therefore accepted in theory and supported by empirical findings.

There is admittedly a distinction between covert and overt hate speech in terms of the likelihood of the victim’s recognition of the speech as hate speech. Take, for example, the following statements ‘You Muslim scum are all terrorist bastards, get out of my country’ and an article in a newspaper claiming that:

\textit{ISIL [a reference to the so-called Islamic State of Iraq and the Levant] is only one manifestation of a global extremism that existed before ISIL and will continue after it and has a very large pool of potential recruits among the new populations of Europe. To diminish that pool, European governments should avoid unnecessary policies (such as policing swimwear) that exacerbate unnecessary grievances and focus instead on those necessary policies - slowing Muslim immigration, carrying out proper}


\textsuperscript{590} Ibid 782,787–789.

\textsuperscript{591} Aisha Phoenix, ‘Somali Young Women and Hierarchies of Belonging’ (2011) 19 YOUNG 313, 322.


vetting of those who arrive and expelling anyone who preaches hatred - whether they cause grievances or not.594

 Scholars who have studied Islamophobia would note that both statements draw on tropes that regard Muslims (all of them, not just a few extremists) as potential terrorists.595 They treat Muslims as uniquely dangerous and prone to violence; because of this, they should be surveilled and excluded from the rest of civilised society.596 Both claim that those Muslims are morally inferior to us. The difference is that the article delivers this message through sonorous prose and moderate language which may be mistaken for legitimate political debate, and the message of inferiority it carries may escape recognition by at least some British Muslims.597

 Further, the more carefully worded and moderate the language is, the less likely it is that the victims will recognise it (and be silenced by it), and it is possible to conceive of a dog whistle with a frequency so high pitched that only a small subset of people with some special knowledge would be able to decode it. Heinze’s example of German far-right groups using ‘87+1’ as a code for ‘Heil Hitler’ is illustrative.598

 However, the question of reach is relevant here. A racial insult directed at someone on the street is heard by and can silence that victim alone, but DNS appearing in the pages of a national newspaper (or on its website) will be read by far more members of the victim group. We are therefore comparing a higher likelihood of recognition by a much smaller number of people to a (comparatively) lower likelihood of recognition by a much larger group.

 It is this reach and influence that gives the press a greater capacity to silence, what Rowbottom refers to as ‘communicative power’.599 The press can produce and reproduce at scale and amplify topics, a phenomenon often described in media effects

---

595 Allen (n 27) 184.
596 ibid.
597 Maitra and McGowan (n 461) 233.
598 Heinze, Hate Speech and Democratic Citizenship (n 5) 146.
literature as ‘agenda-setting’. Sonja West refers to them as ‘repeat players’ because of their unique ability to widely and frequently disseminate content to the public. This is what distinguishes the press from lone or even influential hate speakers. It means that DNS by the press will be read, heard, and viewed much more frequently by many more members of the minority communities than overt hate speech from lone hate speakers, and therefore has the potential to silence at a greater scale.

6. Conclusion

This examination of direct harms has revealed that there is evidence that overt targeted hate speech causes psychological harm and severe emotional distress. However, there is no empirical evidence linking diffuse covert hate speech with these harms. The takeaway from this is that psychological harm and severe emotional distress are largely limited to overt hate speech used in interpersonal contexts. Such hate speech, if committed by individual speakers, is already restricted by racial and religiously aggravated public order offences. Similarly, the press code contains restrictions on targeted hate speech that limits pejorative and prejudicial references to individuals based on protected characteristics. The limitation of such regulation to individuals is justified considering the absence of empirical evidence linking DNS to psychological harm and severe emotional distress. It cannot be claimed that either the diffuse hate speech of lone hate speakers or DNS by the press cause these harms.

On the other hand, it has been noted that targeted overt hate speech amounts to misrecognition. It lowers the self-esteem of members of minority groups and silences them. I have also shown that there is similar evidence that DNS by the press also has such a silencing effect—and does so at a greater scale. Therefore, the fact that DNS causes the same harm (silencing) at a greater scale than overt hate speech by a lone hate speaker justifies imposing broader restrictions on the former. Optimal balancing here requires scaling up the constraints on the press to catch the speech (DNS) causing this silencing. In the next chapter, I engage in a similar analysis of the indirect harms caused by hate speech.

Chapter 5: The Indirect Harms of Dangerous Negative Stereotypes

1. Introduction

In the previous chapter, I explored the various direct harms caused by hate speech. Here I engage in a similar evaluation of indirect harms. Various scholars have examined these indirect harms including Waldron, Tsesis, Brown, and Parekh, among others. Specifically, I look at the role played by hate speech in creating a climate of hatred and undermining the assurance of security and dignity of vulnerable minority groups. I will examine evidence in support of each of these claims and consider whether there is evidence of similar quality suggesting that DNS by the press cause them at a greater scale or intensity.

Section 2 sets out the various indirect harms that are caused by overt hate speech and the empirical evidence in support of them. Section 3 examines whether there is similar evidence linking DNS published by the press with these harms. It will be noted that the media has a unique capacity to cause these harms due to powerful priming and cultivating effects.

2. The indirect harms of overt hate speech

Recall that hate speech can harm indirectly when a person (A) communicates a message which is heard or read by another person (B) who then engages in conduct that harms a victim (C).\(^{602}\) B’s presence between the cause (A’s speech) and effect (harm to C) means that there is a less proximal connection between speech and harm than the direct harms just explored. B’s presence introduces a set of intervening variables that complicate the chain of causation. As such, these claims of indirect harm are generally regarded as more controversial in the hate speech literature.\(^{603}\) However, as with the direct harms, we only require evidence demonstrating that hate speech is a contributing factor to their occurrence.

---

\(^{602}\) Schauer, ‘The Phenomenology of Speech and Harm’ (n 476) 642.
\(^{603}\) Brown, Hate Speech Law (n 55) 52.
a. Violence and discrimination: contributing towards a climate of hatred

One of the commonly cited rationales for restricting hate speech is that it will lead to acts of discrimination and violence against minority groups. We have seen that one of the ways of conceptualising the relationship between hate speech and harm is deterministic. Proponents of this deterministic model argue that hate speech ought to be limited only when it amounts to incitement to specific acts of discrimination and violence, something akin to the fighting words doctrine of the US Constitution.

However, we have also seen that hate speech legislation found in Europe, Australia and Canada, to name a few, is based on a less proximal connection between speech and harm, including acts of discrimination and violence. The view is that these harms are facilitated by the creation of what Brown terms a ‘climate of hatred’ which he defines as the ‘widespread attitude of hatred that is partly the product of an accumulation of hate speech, and that is associated with an increased chance of acts of discrimination and violence’. Violence and discrimination are much easier to justify where certain groups have been dehumanised. There are a plethora of examples from human history which demonstrate this. This was most poignantly expressed by the Canadian Supreme Court in Keegstra which noted:

*The repetition of the loathsome messages of Nazi propaganda led in cruel and rapid succession from the breaking of the shop windows of Jewish merchants to the dispossession of the Jews from their property and their professions, to the establishment of concentration camps and gas chambers. The genocidal horrors of the Holocaust were made possible by the deliberate incitement of hatred against the Jewish and other minority peoples.*

Sottiaux points out that similar reasoning has been embraced by the ECtHR in Féret and Le Pen. These were both cases of Islamophobic hate speech where the court found that the applicant’s statements were ‘capable of arousing particularly among the less informed members of the public, feelings of distrust, rejection or hatred towards

---

604 Goodall (n 17) 220; Brown, *Hate Speech Law* (n 55) 26.
607 Sottiaux (n 316) 48.
foreigners and instilling ‘feelings of rejection and hostility against the targeted community’. The courts here and in other cases recognise that hate speech is the ‘slow-acting poison’ that contributes to the creation of a climate of hatred that lays the groundwork for such atrocities. Parekh makes a similar observation noting that hate speech ‘encourages a climate in which, over time, some groups come to be demonized, and their discriminatory treatment becomes accepted as normal’.

The contribution made by hate speech is what Feinberg would term as an ‘accumulated harm’ akin to environmental pollution, where the harm (degradation of the environment) cannot be traced to a single instance of pollution but is rather a predictable cumulative effect of all the pollution. Similar concerns about the creation of a climate of hatred which would beget violence in the long-term informed Britain’s promulgation of the stirring up offences. The approach in Europe and the UK then is an attempt at intervening earlier in the causal chain between speech and acts of discrimination and/or violence.

By way of evidence for this progression from speech to violence, scholars such as Anthony Cortese, Delgado and Waldron set out the role played by hate speech in various historical genocidal campaigns. In the legal literature on hate speech, Tsesis’s analysis is particularly comprehensive. He systematically reviews historical campaigns of discrimination and mass murder, from the slave trade to the Holocaust and the Rwandan genocide. He also examines contemporary cases of violence at lower scales, such as the campaign of ethnic cleansing in Sudan and post-election violence in Kenya, finding that all were preceded by the dehumanisation of the affected

---

610 This is a phrase used by Waldron, *The Harm in Hate Speech* (n 14) 4.
611 Parekh, ‘Is There a Case for Banning Hate Speech?’ (n 569) 45.
612 Feinberg, *The Moral Limits of the Criminal Law* (n 326) 225–229; Simpson (n 9) 707.
groups through hate speech. The view that hate speech is a pre-condition to large-scale discrimination and violence is widely ascribed to in atrocity scholarship.

Allport sheds some light on how speech can progress to discrimination and violence using a five-point scale. Prejudice is first expressed as antilocution (hate speech), then avoidance (the ingroup avoids interacting with the outgroup), discrimination (denial of equal treatment, e.g. housing, services), physical attacks (threats, vandalism, violence), and finally, extermination (genocide).

Allport notes that the scale is not ‘mathematically-constructed’, but generally, ‘activity on one level makes the transition to a more intense level easier’. Parekh refers to similar processes of escalation. Hate speech is the first point on the scale, a part contributor to creating a climate of hatred that makes progression to these latter stages more likely.

Heinze objects to this framing, arguing that unlike the Weimar Republic or Rwanda, stable western democracies such as the UK have sufficiently robust political structures to prevent mass discrimination and violence against minority groups. However, discrimination and violence at a much lower scale are still substantial evils, particularly in light of the well-documented rise in hate crime in the UK in the last ten years. Legislators may therefore argue that they need not wait for mass discrimination and/or violence to eliminate or reduce at least one of the factors contributing to the creation and sustenance of the climate that feeds these hate crimes.

---

617 Allport is a social psychologist whose work focuses on prejudice that produces ‘exclusion and violence’ John F Dovidio, Peter Samuel Glick and Laurie A Rudman, On the Nature of Prejudice: Fifty Years after Allport (Blackwell Publishing 2005) 10.
619 ibid 15.
620 Parekh, ‘Is There a Case for Banning Hate Speech?’ (n 569) 45.
621 Kopytowska and Baider (n 57) 138.
624 Brown, Hate Speech Law (n 55) 67.
b. Undermining security and dignity

It is also not just a question of whether people are discriminated against or attacked but whether they are afraid of this occurring. As Waldron notes, how a society looks provides an indication to minority groups of how they are ‘likely to be treated, for example by the hundreds of thousands of strangers they encounter’ in daily life. Minorities who live in a society where a substantial number of people hold them in contempt, partly as a result of a climate of hatred created by hate speech and where such speech is a regular feature of the visible environment, may modify their behaviour and avoid certain places for fear of being abused, shunned or attacked. He argues that by excising hate speech out of the visible environment, hate speech laws provide minorities with the assurance and felt sense of security they need to live their lives.

Finally, Waldron argues that the absence of discrimination and violence is a limited view of public order. He notes that hate speech laws are meant to uphold people’s dignity. The term dignity is a malleable concept, defined and used in different ways by legal scholars on both sides of the hate speech debates. Oppositionists such as Ronald Dworkin and Kent Greenawalt have argued that dignity is violated by paternalistic restrictions on speech, whereas prohibitionists such as R. George Wright and Steven Heyman claim that hate speech undermines dignity. Heyman’s claim is based on the direct impact of the speech on the target which he argues ‘constitutes a fundamental attack on their right to personal dignity’. Heyman’s conception of dignity is akin to the direct harm of misrecognition considered in the previous chapter. Waldron’s account of dignity differs from these, he notes that

---

625 Waldron, *The Harm in Hate Speech* (n 14) 82.
629 Dworkin, *Freedom’s Law* (n 455) 238.
630 Greenawalt (n 287) 28, 33–34.
632 Steven J Heyman, *Free Speech and Human Dignity* (Yale University Press 2008).
633 Heyman (n 555) 166.
philosophically dignity may well be a Kantian notion of inherent moral worth, but the dignity protected by hate speech laws is:

one’s status as an ordinary member of society in good standing, entitled to the same liberties, protections, and powers that everyone else has—and it generates demands for recognition and treatment which accords this status.

He views dignity as social standing, the esteem by which a person is held by others in society. This view is rooted in the ancient roman concept of dignitas meaning ‘status honour and respect’; while in that society, the rank was a preserve of the elite, today it is ‘apportioned out in equal measure, insofar as we all occupy an equivalent rank, as members of a shared moral community’.

Christopher McCrudden notes that while the ‘use of dignity in human rights texts’ reflects ‘different strands of metaphysical and philosophical thinking’, the ‘minimum core’ includes both dignity as inherent moral worth and as social standing to be ‘recognised and respected by others’. In some ways, it is the twin to the harm of misrecognition. Whereas targeted hate speech tarnishes one’s standing in their own eyes, diffuse hate speech tarnishes it amongst their peers, undermining their position as their equal in rank and status and perpetuating racial/religious social hierarchies.

Waldron’s account has been understood and applied in different ways, and there is admittedly some ambiguity as to whether he believes hate speech merely causes or also constitutes harm. However, what is key to his account is this more expansive notion of public order that is not limited to freedom from violence and discrimination but also people’s ‘right to be treated as members of societies with a status and acceptance that enables them to participate confidently in the ordinary routines and

---

635 Waldron, The Harm in Hate Speech (n 14) 219–220.
638 Barendt, ‘What Is the Harm of Hate Speech?’ (n 289).
transactions of everyday social life’, it is this status that enables the maintenance of cohesion and public order in a multicultural, pluralist society.\(^{639}\)

Further, lowering someone’s social standing can infringe on other interests. Some argue that it can cause silencing; if hate speech causes others to believe a certain group belongs to a lower caste,\(^{640}\) they will not consider them their equals in democratic deliberation, and the speech of the targeted group’s members will ‘fall on deaf ears’.\(^{641}\)

Therefore, there are at least three reasons to restrict overt hate speech; to provide objective security by preventing the creation of a climate of hatred that begets violence in the long term, to provide subjective feelings of security to minorities which allows them to live their lives without fear and to prevent their social standing from being undermined or lowered.

Notably, all these claims draw force from and turn on a single empirical claim; that hate speech can increase people’s prejudice towards minority groups. If hate speech had no effect on people’s attitudes towards minorities, these claims become less persuasive. Hate speech would not feed a climate of hatred, minorities may have no reason to feel threatened, and their social standing would remain intact. This empirical claim is one which is contested. Simpson argues that it is not clear what role hate speech plays in the ‘complex causal-systems—involving a mixture of institutional, material, economic and discursive factors’ which inform the creation of ‘identity-based social hierarchies’.\(^{642}\) To Simpson, hate speech may either be a contributing factor or it might be a ‘downstream consequence’ of these other factors.\(^{643}\) We must then have evidence demonstrating that overt hate speech contributes to fostering prejudice.

Matsuda argues that such evidence does exist, citing a study demonstrating that exposure to overt hate speech in the form of racial slurs caused listeners to negatively

\(^{639}\) Waldron, The Harm in Hate Speech (n 14) 219.
\(^{640}\) I borrow this analogy from Simpson (n 9) 709.
\(^{641}\) Brown, Hate Speech Law (n 55) 198.
\(^{642}\) Simpson (n 9) 722–723.
\(^{643}\) ibid 707, 723.
evaluate those it was directed at.\(^{644}\) There have been numerous subsequent studies since. Here it seems that—unlike those examining psychological harms—researchers have managed to navigate the ethical constraints that militate against intentionally exposing participants to hate speech. One such study by Fasoli and others found that homophobic epithets had a dehumanising effect and caused people to physically distance themselves from gay men.\(^{645}\) Similarly, Soral and others examined the effect of exposing participants to overt hate speech directed against Jews, Muslims, Ukrainians and the LGBT community, finding that repeated exposure to hate speech led to ‘greater desensitisation to such statements’ and an ‘increase in outgroup prejudice’ against these groups.\(^{646}\)

Furthermore, outside of laboratory conditions, researchers have been able to demonstrate an association between offline and online hate speech and an increase in hate crimes against the targeted group in stable democracies.\(^{647}\) The current state of the empirical evidence seems to have convinced the Law Commission which concluded in its final report on the review of British hate crime and hate speech laws that: ‘there is extensive evidence that the propagation of inflammatory hate material can lead to violence and disorder’.\(^{648}\)

3. Press hate speech and indirect harms

The question then is whether DNS by the press can also foster prejudicial attitudes towards minorities. The Council of Europe certainly believes this to be the case noting


\(^{646}\) Wiktor Soral, Bilewicz Michal and Winiewsk Mikołaj, ‘Exposure to Hate Speech Increases Prejudice through Desensitization’ (2018) 44 Aggress Behav 136, 138,144.


\(^{648}\) ‘Hate Crime Laws: Final Report’ (n 165) 392.
that hate speech ‘may have a greater and more damaging impact when disseminated through the media’.\textsuperscript{649} So did the Leveson inquiry, which questioned:

\textit{Whether articles unfairly representing Muslims in a negative light are appropriate in a mature democracy which respects both freedom of expression and the right of individuals not to face discrimination}.\textsuperscript{650}

I argue that these views are correct for two reasons. Firstly, scholars who claim that overt hate speech causes these harms are in fact concerned with the perpetuation of DNS. Secondly, a substantial—and growing—body of research into media effects on priming and cultivation shows that exposure to press hate speech is linked with an increase in prejudicial attitudes towards minorities.

\textbf{a. DNS as a source of indirect harms}

Turning to the first point, scholars often claim that these harms are caused by speech that amounts to ‘group defamation’. Parekh defines group defamation as the making of ‘public untruthful and damaging remarks’ about a group which lowers ‘them in their own and other’s eyes’.\textsuperscript{651} Waldron adopts a similar definition referring to speech that attacks a person’s status and ‘reputation as a citizen or member of society in good standing’.\textsuperscript{652}

On closer inspection, most of the instances of hate speech that they cite are really examples of DNS. Parekh refers to statements such as ‘all Jews are secretive, greedy, vindictive and conspiratorial’ or ‘all blacks are stupid, unruly, licentious and unreliable’.\textsuperscript{653} He also mentions an article in a Danish newspaper about the Danish cartoon affair which portrayed Muslims as backward, violent, and intent on introducing ‘Islamic law’ in Denmark.\textsuperscript{654} To Waldron, speech that amounts to ‘slow-acting poison’ includes a poster claiming Muslims are ‘all Osama’ and another stuck on the wall of a mosque reading ‘Jihad Central’.\textsuperscript{655} Tsesis writes that ‘racist instigation ascribes

\begin{thebibliography}{9}
\item Recommendation No. R (97) 20 of the Committee of Ministers to Member States on ‘Hate Speech’.
\item Leveson (n 2) 671.
\item Waldron, \textit{The Harm in Hate Speech} (n 14) 313.
\item \textit{Ibid} 47.
\item Parekh, \textit{Rethinking Multiculturalism} (n 57) 313.
\item Waldron, \textit{The Harm in Hate Speech} (n 14) 1.
\end{thebibliography}
undesirable traits to disparages groups—greediness to Jews lasciviousness to Blacks, laziness to Mexicans—in order to diminish their political and social standing’.  

You will note that these are examples of anti-black, antisemitic and Islamophobic stereotypes. They are a re-circulation of age-old tropes that have been repeatedly used against these groups, the unintelligent Black person, the ‘devious and cunning’ Jew, and the violent Muslim. Brown is therefore correct when he notes that scholars arguing about regulating group defamation are in fact advocating for ‘regulating the use of negative stereotypes, especially in the media’. It is the perpetuation of these DNS that is central to their understanding of what constitutes hate speech and the means by which groups are defamed, and hatred against them is stirred up. If this is the case, there is no reason to believe that this is only caused by overt hate speech; ‘don’t trust these kikes, the holocaust is a fucking lie’ but not when expressed through moderate language; ‘We will decriminalise Holocaust denial and enter a correct teaching of the history of the Third Reich’. Both perpetuate DNS that Jews are engaged in a grand conspiracy to falsify the Holocaust for their personal advantage. The content of the speech matters more than its form. Heinze is justified when he argues that if one believes that hate speech causes these indirect harms, limiting their focus to only the ‘most overtly violent or extreme expressions of hatred’ is ‘patently inadequate’.

b. Press hate speech: Priming and cultivating prejudice

We also have empirical evidence of—at least—similar quality demonstrating that DNS also contribute to indirect harms and do so to a greater degree. In fact, Brown notes

---

656 Tsesis (n 615) 505.
659 Allen (n 27).
660 Brown, Hate Speech Law (n 55) 23,178-180.
661 This example is from the Handbook for the Practical Use of the IHRA Working Definition of Antisemitism (n 88) 13.
662 ibid 12.
663 Heinze, Hate Speech and Democratic Citizenship (n 5) 165.
that the bulk of the social science research studying the link between speech and indirect harms is in communication studies and media effects.\textsuperscript{664}

Similar to the evidence linking overt hate speech to these harms, media effects are conditional and contingent. The early deterministic ‘hypodermic needle’ media effects theories, which suggested that media messages were beamed into the minds of the audience and directly accepted have given way to more sophisticated probabilistic models recognising that media effects are complex.\textsuperscript{665} They are enhanced or reduced by a plethora of factors, including individual differences between consumers.

Many variables, such as personality traits, temperament, mood and values, impact both media choice and influence.\textsuperscript{666} This is pertinent as people can only be influenced by what they consume—including hate speech.\textsuperscript{667} Further, cognitive biases such as motivated reasoning mean that people tend to process information in ways that conform to their pre-existing beliefs.\textsuperscript{668} A person with pro-immigration views is less likely to seek out and have their attitudes and views influenced by an article advocating for the imposition of strict anti-immigration measures. Conversely, those whose values align with these policies are more likely to seek out such articles and have their attitudes reinforced by them.

This certainly is not unique to the media; these are claims about human cognition that equally apply to the processing of hate speech from other speakers. The point is simply that when evaluating the evidence of media effects, we should be cognisant of these variables that determine its influence and avoid overselling or catastrophising the media’s capacity to harm. Moreover, such claims are wholly unnecessary because to justify regulation, one need only demonstrate that DNS by the press also contribute

\textsuperscript{664} Brown, Hate Speech Law (n 55) 69.
\textsuperscript{667} Valkenburg, Peter and Walther (n 436) 324.
towards the creation of a climate of hatred and do so to a greater degree than hate speech by a lone hate speaker. Such evidence is certainly forthcoming.

c. The effectiveness of moderate language

Moderate language is more effective at persuading an audience to adopt bigoted attitudes towards minorities than overt hate speech. I noted earlier that because of a general liberalisation in social attitudes, overt displays of bigotry, ‘Gypsies are scum,’ are off-putting even to audiences who hold intolerant views. It may even produce sympathy ‘look, I don’t have much time for Gypsies either, but you’ve gone too far’. Conversely, covert hate speech can have the opposite effect. Its moderation means that even those with more liberal attitudes may conclude that the speech falls within the Overton Window, meaning they may give the message due consideration instead of outrightly rejecting it: ‘I disagree with you on this, but you may have a point’. Equally, differentiation within press hate speech also has a similar effect. Audiences with greater outgroup prejudice may find the ‘open ridicule’ of minorities in the tabloids more acceptable, whereas those with more tolerant beliefs are more susceptible to the more subtle and palatable messaging in the serious press. Oppositionists such as Heinze and prohibitionists including Sorial and Gelber note that this makes ‘sophisticated racism’ more dangerous as it is more likely to shape the views of the audience than crude and ‘grosser’ overt hate speech.

The fact that some sections of the British press do engage in hate speech, that this hate speech is primarily through perpetuating DNS, that it is a form of covert hate speech and the social, historical and legal factors that have influenced these changes have already been set out in earlier and I will not repeat that here. Instead, I only focus on whether DNS have an effect on attitudes and/or behaviour towards these groups.

---

669 See Chapter 2.
670 ‘Hate Crime Laws a Consultation Paper’ (n 161) 34.
671 Elizabeth Poole, Reporting Islam: Media Representations of British Muslims (Continuum-3PL 2002) 83.
672 Sarah Sorial, ‘Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement’ (2015) 34 Law and Philosophy 299, 301; Gelber, Speaking Back (n 134) 79; Heinze, Hate Speech and Democratic Citizenship (n 5) 151.
673 See Chapter 1.
Stereotypes about social groups come from various sources in one’s social environment, such as peers, parents and culture. The institutional press is a principal source of stereotypes for various reasons. Firstly, they widely and frequently disseminate content to the public. Secondly, people are more dependent on the media for information about groups that they lack ‘personal or vicarious experience with’. In this respect, they already produce greater harm than a lone hate speaker.

Media stereotyping need not be intentional as it can be perpetuated through what media effects scholars refer to as ‘framing’. Framing refers to the angle or perspectives chosen by the media to report on a story. It is argued that the choice to emphasise or de-emphasise certain aspects impacts the audience’s perceptions. Stories can be framed in ways that perpetuate stereotypes. For example, when reporting on a refugee crisis, the story could focus on the plight of the newcomers, individualise them and provide information about their background, e.g. a nurse fleeing a conflict zone, or it could focus on the ‘threat’ the refugees pose to the host country and/or the strain they place on resources and public services. The latter perpetuates the stereotype that refugees are threats, and this could be encoded in the memories of the audience. Threat framings are also likely to be more vividly remembered and more widely circulated as people tend to pay more attention to negative news. This also creates economic incentives to pursue these stereotyping portrayals.

But the fact that most of our knowledge about the content of stereotypes about minority groups comes from the media does not tell us that they contribute to the creation of a climate of hatred, as knowledge is not the same thing as personal beliefs or behaviour. Patricia Devine argues that personal beliefs ‘may or may not be congruent’ with our knowledge about the stereotype. Many will be familiar with the content of the various examples of Antisemitic, Anti-black and Islamophobic tropes set out in this thesis. However, they will likely reject these stereotypes because of their egalitarian values and not let them form part of their personal beliefs or inform their behaviour towards

---

674 Nelson (n 57) 13.
675 Gavin (n 39) 832.
676 James Curran, Media and Society (5th edn, Bloomsbury Academic 2010) 279.
677 Dixon and Maddox (n 87); Florian Arendt, ‘Dose-Dependent Media Priming Effects of Stereotypic Newspaper Articles on Implicit and Explicit Stereotypes’ (2013) 63 Journal of Communication 830, 831.
678 Valkenburg, Peter and Walther (n 436) 320.
Black people, Jews or Muslims. Therefore, something more must be said to bridge the gap between knowledge on the one hand and attitudes and behaviour on the other.

I. The power of media stereotyping: System 1 and System 2 thinking

Media effects researchers have relied on literature in behavioural economics, as well as cognitive and social psychology to bridge this gap. They argue that media stereotypes activate cognitive shortcuts that lead audiences to develop attitudes and/or behaviour that conform to stereotypes. I will first briefly describe the nature of these cognitive shortcuts/heuristics and then set out the evidence that supports their claims. Scholarship examining information processing has distinguished between two distinct methods of decision-making. The first is what Daniel Kahneman refers to as ‘System 1’ decisions. These are automatic and quick ‘with little or no effort and no sense of voluntary control’.680 The other is System 2 which is slower and is used for ‘mental activities that demand it, including complex computations’.681

System 1 is the system we use most often as ‘people have limited time, information, and cognitive resources’.682 It allows us to come to ‘quick and dirty’ decisions that rely on heuristics.683 When people are required to make judgements about something, they often retrieve the most available information from their memories that are relevant to that thing. This is referred to as an ‘availability heuristic’.684 Availability is predicated on several factors, including salience, vividness and recency of the information.685 The idea here is that where an individual is required to make a judgment either about a social group or issues pertaining to them, they mostly use System 1 decisions, relying on the stereotypes encoded in their memory as an availability heuristic to inform those decisions.686

The media is not only a source of much of the original ‘encoding’ of these stereotypes in people’s memories; they also increase their availability whenever they publish

---

681 ibid 25.
682 Andersen Jones and Grow Sun (n 668) 438.
683 Kahneman (n 680) 25.
685 ibid.
content containing these stereotypes as this causes the ‘activation of those constructs in memory’. This is referred to as priming.687 Dixon notes that activation through media priming informs the consumers ‘subsequent (and typically immediate) cognitive, attitudinal, or behavioural decisions’.688 By informing here, they mean that after priming the consumer may process information, adopt attitudes, make judgements and exhibit behaviour which is prejudicial towards the stereotyped group.689 Since this process is automatic it applies equally to low and high prejudice individuals.

Various studies have been able to demonstrate this effect. These studies exposed participants to media content containing stereotypes and then examined post-exposure attitudes and behaviour towards these groups. Some found links between media stereotypes and subsequent attitudes but not in behaviour. One such study examined the effect of exposure to stereotyping of immigrants as criminals and threats to the host country’s culture and economy, noting that such exposure led to an increase in contempt towards the outgroup.690 However, it did not result in active support for policies that harmed these groups.691

Others linked media stereotypes to both attitudes and behaviour. A study found that the exposure to a news story containing the ‘black criminal stereotype’692 resulted in ‘harsher culpability judgments of subsequent criminal suspects’ and increased support for policies that disproportionality affected the stereotyped groups, such as the death penalty.693 These findings have been replicated in subsequent studies.694 Another study found that participants exposed to media stereotypes about African Americans were more likely to demonstrate prejudicial behaviour towards them; in this case, they were more likely to find them guilty in a fictitious assault case.695 Another study found

688 Dixon and Maddox (n 87) 1556; Dixon (n 57) 249.
689 Ibid.
691 Ibid 832.
692 This is the well-known trope depicting black people as predisposed to committing—often violent—crimes or which presents them as ‘criminal predators’ see generally Kelly Welch, ‘Black Criminal Stereotypes and Racial Profiling’ (2007) 23 Journal of Contemporary Criminal Justice 276.
693 Dixon and Maddox (n 87) 1562–1564; Valkenburg, Peter a
a correlation between media stereotypes portraying Muslims as terrorists and subsequent negative attitudes (perception of Muslims as threats) and behaviour (support for policies that harmed them, such as denying Muslims the right to vote and surveilling them).696

Notably, many of these experiments tested short-term effects but the researchers hypothesise that the effects were more long-term, and these long-term effects would be even more pronounced.697 The idea is that ‘whereas priming suggests short-term activation of a cognitive linkage used to make an immediate subsequent judgment, the repeated activation of such a link can increase its accessibility’ making this process more automatic and therefore increasing the reliance on these stereotypes as heuristics to inform attitudes and behaviour.698 This is particularly problematic as we have already seen that these decisions are largely automatic, and therefore, these attitudes and behaviours manifest ‘even without conscious awareness’.699 These long-term effects have been tested and supported by various studies.700 One study found that over a 20-year period, public attitudes towards immigration in Britain tracked negative coverage in the tabloids.701 The scale of this influence alone cannot be matched by lone hate speakers.

As a response to this, it is argued that we use System 2 to control our impulses and intuitions generated by System 1.702 A question arises as to whether System 1 processes can be bypassed by deliberative System 2 decision-making to inhibit the ‘automatically activated stereotypes’.703 Devine makes exactly this claim noting that low-prejudice individuals can use System 2 to consciously inhibit prejudiced System 1 beliefs.704 Studies that test this demonstrate that high-prejudice people were more

697 Dixon (n 57) 249.
698 ibid.
700 Anita Atwell Seate and others, ‘Cultivating Intergroup Emotions: An Intergroup Threat Theory Approach’ (2018) 21 Mass Communication and Society 178; Saleem and others (n 696) 844–845; Dixon (n 57) 249.
702 Kahneman (n 680) 26.
703 Devine (n 679) 7.
704 ibid.
likely to endorse and act consistently with the stereotype, whereas low-prejudice people often 'censored and inhibited' the automatically activated stereotype and 'replaced it with thoughts that expressed their non-prejudiced values'.\textsuperscript{705} Egalitarian values are therefore an important mediator which potentially weakens the claim that DNS by the press can contribute to a change in attitudes and, therefore, towards the creation of a climate of hatred.

However, several factors also weaken the impact of this mediator. Firstly, it has little to no impact on high-prejudice individuals. While the press does not change their attitudes towards minorities—as they are already prejudiced—their attitudes can be further entrenched through the cultivating effects of the repeated activation of stereotypes via media priming. One concern is that these entrenched beliefs desensitise them towards more extreme content that could further radicalise them.\textsuperscript{706} In fact, one of the government’s justifications for criminalising overt hate speech is that ‘even those already holding racist views can be incited into further hatred’.\textsuperscript{707} Another concern is that entrenched attitudes are more impervious to persuasion, and where a climate of hatred already exists, DNS make it harder to uncreate it.

Secondly, there are reasons to question the impact of the mediator for even low-prejudice individuals. Devine notes that one reason is that ‘stereotypes have a longer history of activation and are therefore likely to be more accessible than are personal beliefs’.\textsuperscript{708} Recall also that System 1 processes are automatic, whereas System 2 must be deliberately activated. In order to circumvent this automatic process (of using stereotypes as heuristics) and act/respond in a non-prejudiced manner, low-prejudice individuals must consciously monitor when they are activated, inhibit them and replace them with their non-prejudiced beliefs.\textsuperscript{709} This requires significant cognitive resources leading one scholar to note that ‘unless the perceiver has the motivation to be accurate in making judgments, automatic stereotype processing will be the preferred mode in

\begin{flushleft}
\textsuperscript{705} ibid 14.  \\
\textsuperscript{706} Soral, Michał and Mikołaj (n 646).  \\
\textsuperscript{707} ‘Hate Crime Laws: Final Report’ (n 165) 37.  \\
\textsuperscript{708} Devine (n 679) 6.  \\
\textsuperscript{709} ibid 7.
\end{flushleft}
comparison to deliberate, individuated processing'.

Therefore, even for low-prejudice individuals, non-prejudiced responses may be the exception, not the norm.

Thirdly, an important practical factor is that intragroup contact may impact the motivation to use System 2 to inhibit System 1 stereotyping heuristics. We also know that System 2 is more likely to be activated where we are faced with data that is contrary to our expectations. Therefore, those who have members of minority groups in their social circles will be more motivated to activate System 2 when they read an article containing DNS: ‘that can’t be right Muslims are not terrorists. I have many Muslim friends’. As evidence for this point, Elizabeth Poole’s study on the impact of negative press coverage of British Muslims found that non-Muslims who had no close contact with Muslims were more likely to have their attitudes changed and shaped by media portrayals. Other studies have made similar findings, noting the impact of this ‘cultural proximity’ on mediating media stereotyping. A potential response therefore, is that the real problem here is the lack of outgroup contact, and we should look at ways of ensuring greater integration between communities instead of restricting speech. Brown pushes back against this, arguing that such policies are not mutually exclusive; rather, ‘what this shows is that society is likely to need all of these other policies alongside regulation of negative stereotypes in the media.’

Brown’s view is supported by the reality of population differences (86% of the population in Britain is White), which make meaningful intragroup contact difficult to achieve with even the most far-reaching integration policies. Meanwhile, the data that we currently have from surveys indicate that 35% of White Brits have no ethnic minority friends, and 40% have ‘never had close contact with Muslims’. This means that as matters stand, a significant part of the population, including low-prejudice

710 ibid 7.
711 Poole (n 671) 213.
713 Brown, Hate Speech Law (n 55) 70.
715 ‘One Third of White Britons Don’t Have Any Friends from an Ethnic Minority Background | YouGov’ <https://yougov.co.uk/topics/politics/articles-reports/2018/05/03/one-third-white-britons-dont-have-any-friends-ethn> accessed 21 April 2022.
716 ‘Islamophobia Defined: The Inquiry into a Working Definition of Islamophobia’ (n 183) 9.
individuals, may not even have the motivation to sum up the cognitive resources and engage System 2 to counter stereotyping heuristics.

Fourthly, intragroup contact alone may be insufficient as rational decision-making is bounded by a lack of subject knowledge. Poole’s study shows that this impacts people’s ability to counter stereotyping media depictions. It found that a common response by those who have Muslim friends but who knew little about Islam was not to reject media stereotypes but rather detach them from their friends: ‘Islam may promote terrorism, but my friend is a good person’. Rather, rejecting stereotypes requires subject matter knowledge, ‘I am familiar with this chapter of the Quran, and this is incorrect’. Therefore, consumers require motivation from intragroup contact and knowledge about the stereotyped group’s culture/beliefs to avoid being influenced by media stereotypes. This means only a small subset of the population is inoculated from the attitudinal and behavioural effects of press hate speech. Notably, recent surveys suggest that many Brits either know nothing about what Muslims believe or those who think they do are often misinformed.

Even setting aside the evidence of attitudinal changes, there’s a question as to how this content impacts a minority group’s felt sense of security. Recall that the visibility of hate speech in an environment is a cause of subjective insecurity. Press hate speech forms a much larger part of our visible environment than posters and pamphlets by lone hate speakers. Physical copies appear in newsstands, supermarkets, and subway stations with headlines such as ‘Muslim Schools Ban Our Culture’, ‘BBC Puts Muslims Before You’, ‘Brit Kids Forced to Eat Halal School Dinners!’, ‘Muslims Tell Us How to Run Our Schools’. They also form part of the online environment through the online versions of their papers which are also widely

---

718 Poole (n 671) 208.
720 Waldron, The Harm in Hate Speech (n 14) 57.
circulated on popular social media platforms. Further, the phenomena of ‘third person effects’ is relevant here. It is a very well-supported media effects theory which claims that people tend to perceive that the media has a greater effect on other people than it has on themselves. This means that minorities are likely to overestimate the media’s capacity to generate hostile attitudes toward them. If this is the case, press hate speech may create in them a greater fear of being shunned, discriminated against, and attacked than a lone speaker’s hate speech. There is evidence to suggest this is the case.

Finally, some scholars contend that hate speech in the mass media can incentivise hate crimes in quite unexpected ways. Dhammika Dharmapala and Richard McAdams argue that a potential offender (A) must weigh up the various intrinsic utilities in committing a hate crime (B), e.g. thrill-seeking, retaliation, mission offenders against the various disincentives such as potential detection and imprisonment (C). It is assumed that whenever B exceeds C, the crime will be committed. However, A is also motivated by the esteem hate speakers get ‘from those who approve of hate crimes’, including esteem from strangers. The greater the esteem, the greater B becomes, and therefore all else being equal, esteem increases the likelihood of the hate crime being committed. Dharmapala and McAdams note that the ‘number of esteeming strangers is not known with certainty’ and A must estimate the number by relying on some public information. This is where hate speech comes in. Every instance of hate speech is an indication to A that another member of the public

---

724 Community consultation by the APPG on British Muslims indicated that many members of the community perceived the media as the cause of their demonisation ‘Islamophobia Defined: The Inquiry into a Working Definition of Islamophobia’ (n 183) 55.
725 Here I am talking about a category of people who are highly prejudiced and want to commit such crimes.
728 This model has been simplified for purposes of the analogy and the full model is set out in ibid 109.
729 Ibid 108.
approves of their crimes. Therefore, the greater the visibility of hate speech, the more esteemers that A perceives exists and the greater B becomes. This is relevant because of the difference in communicative power and legitimacy between lone hate speakers and the media. Publications in the press are likely to suggest to A that there exists in the general populace a far-greater number of esteemers than, say, a lone extremists post on an internet forum. If stories claiming that a certain racial or religious group is dangerous appears in press outlet with wide circulation, it grants that view added legitimacy and conveys to A that these are widely held beliefs, therefore increasing the likelihood that they will commit a hate crime.

4. Conclusion

There is significant evidence demonstrating that DNS by the press fosters prejudice to a greater degree than overt hate speech by lone hate speakers. Therefore, the various indirect harms that flow from prejudice; contributing to the creation and maintenance of a climate of hatred and the erosion of assurance and dignity, are all caused to a greater degree by the press.

Like the direct harms just explored in the previous chapter, this has implications for broader regulation under the equivalence model. The fact that press hate speech causes the same category of harm to a greater degree means that they should be subject to broader restrictions. These restrictions should restrict the content causing these harms (DNS). Furthermore, the failure to regulate such content undermines the goals of existing legislation. If overt hate speech must be regulated to avoid creating a climate of hatred, then failure to regulate DNS by the press which does so to a greater degree, undermines these goals. Similarly, it is questionable whether bans on overt hate speech provide an assurance of security when such assurance is undermined by the press. Finally, it is dubious whether these bans maintain dignity when the press can, to borrow Heinze’s words, ‘mock and scorn vulnerable targets within the bounds of law’. 731 Such broader regulation is entirely compatible with the libertarian theory and its equivalence model.

731 Heinze, ‘Hate Speech and the Normative Foundations of Regulation’ (n 622) 600.
The next chapter presents a different argument. It sets out how the press is unique as unlike most lone hate speakers, their speech doesn’t just cause harm, it constitutes it.
Chapter 6: Dangerous Negative Stereotypes and Constituting Harm

1. Introduction

This chapter presents a supplementary argument to those who reject or are unsatisfied by the causal claims and the state of the evidence in the social sciences explored in chapters 4 and 5. Here, I argue that unlike overt hate speech by ordinary hate speakers which only cause harm, DNS by the press both cause and constitute harm.

Section 2 first sets out the nature of constituted harms and explains the authority problem which prevents the speech of ordinary hate speakers from constituting harm. Section 3 looks at some responses to the authority problem noting that the answers are unpersuasive and that lone hate speakers will only have the requisite authority for their hate speech to constitute harm in limited circumstances. Section 4 argues that the authority problem is either lessened or entirely absent with respect to the press.

The implication and effect of this will be explored in section 5. It will be argued that constituted harms are best understood as enacting harmful social norms which legitimise and incentivise discrimination and that these harms justify imposing broader hate speech regulation on the press under the equivalence model.

2. Constituting harm and the authority problem

Claims that hate speech constitutes harm are built on J.L Austin’s theory of speech acts. Austin argued that speech doesn’t just say things; it does them.732 The relevant distinction is between an utterance’s perlocutionary effect and its illocutionary force. Perlocutionary effects are the causal effects on the audience from the speech such as persuading, scaring, or insulting them.733 Claims that hate speech causes psychological harm, or persuades others to view its targets as undesirable are all perlocutionary effects; they are harms that ‘occur as a result of the speech’.734 This

---

733 ibid.
means that the direct and indirect harms explored thus far are all perlocutionary effects of hate speech.

The illocutionary force of an utterance on the other hand is the thing done or ‘action constituted’ by the speech, such as warning, ordering, promising etc.\textsuperscript{735} ‘I order you to shut the door’ constitutes the illocutionary act of ordering. Here nothing more needs to be done to order. The utterance itself constitutes the act of ordering. The claim that speech constitutes harm is a claim that the speech itself \textit{is} the harm.\textsuperscript{736}

Scholars such as Rae Langton\textsuperscript{737} and Jennifer Hornsby\textsuperscript{738} have relied on speech act theory to claim that pornography subordinates women.\textsuperscript{739} The harm of subordination as described by Langton includes ranking people as less worthy than others, depriving them of important rights and powers and legitimising discriminatory behaviour towards them.\textsuperscript{740} Pornography is said to both cause and constitute at least two of these; it ranks women as ‘sex objects’ and ‘legitimates sexual violence’ towards them.\textsuperscript{741} The claim is not just that subordination is a perlocutionary effect where pornography persuades others to view and treat women as sex objects or convinces its consumers to discriminate against women. Rather it also constitutes these acts. It does not \textit{just} cause the harm of subordination; \textit{it is} subordination.

Ranking appears to be a dignitarian harm in the Waldronian sense discussed earlier.\textsuperscript{742} Here the ranker’s speech act creates (not just causes) a caste system and amounts to a diminution of the social status of the targeted groups. Simpson puts it best when he notes that minorities are ‘vying to secure a position of esteem and recognition within a wider social and cultural ecosystem, and that this aspiration is what is being impaired’ by the speech.\textsuperscript{743} It also legitimises and paves the way for
subsequent (causal) harms. Langton writes that social hierarchies are how ‘feelings of inferiority and superiority are engendered’ and how ‘indifference to violence against those on the bottom is rationalised and normalized’.\textsuperscript{744} Brown also builds on the under-theorised concept of being deprived of ‘important rights and powers’ by proposing that they should be understood as the deprivation of ‘real access to basic human functioning’ such as health, self-worth, and intellectual development.\textsuperscript{745} Finally, subordination can cause silencing, the speech of those who are deemed to be part of a lower caste is not given due consideration and this deprives the ranked groups from effective participation in the democratic process.\textsuperscript{746} This broader concept of subordination encompasses much of the same harms that have been explored thus far, with the key difference being the causal mechanism. This causal mechanism is what will be explored here. We already know that hate speech can cause subordination. The question is whether it also constitutes it.

Lawrence,\textsuperscript{747} Matsuda,\textsuperscript{748} Langton,\textsuperscript{749} Andrew Altman,\textsuperscript{750} Mary Kate McGowan\textsuperscript{751} and Maitra\textsuperscript{752} claim that it does. To demonstrate an illocutionary act of hate speech, Langton uses the example of a legislator in apartheid South Africa who, when enacting legislation depriving Black South Africans of the right to vote, declares, ‘Blacks are not permitted to vote’.\textsuperscript{753} Here the speech has illocutionary force; it ranks blacks as worth less than whites as a matter of law, deprives them of an important right (to vote), and normalises their discrimination.\textsuperscript{754} The speech is the harm and is in addition to any other harmful perlocutionary effects that may causally flow from it, e.g. Black people avoiding areas with white people, the perpetuation of racist attitudes amongst White South Africans etc.

\textsuperscript{744} Langton (n 742) 124.
\textsuperscript{745} Brown, \textit{Hate Speech Law} (n 55) 81.
\textsuperscript{746} ibid 84–85.
\textsuperscript{747} Lawrence (n 12).
\textsuperscript{748} Matsuda (n 510) 2357.
\textsuperscript{749} Langton (n 742).
\textsuperscript{750} Andrew Altman, ‘Liberalism and Campus Hate Speech: A Philosophical Examination’ (1993) 103 Ethics 302.
\textsuperscript{751} McGowan (n 288).
\textsuperscript{752} Ishani Maitra, \textit{Subordinating Speech 1} (Oxford University Press 2012).
\textsuperscript{753} Langton (n 737) 302.
\textsuperscript{754} Langton (n 742) 303.
Another example is a general manager of a private company declaring, ‘we aren’t going to be hiring any more Jews’ or a college president stating, ‘no more women in this department’. These are all instances of ‘authoritative illocutions’.\textsuperscript{755} According to Maciej Witek, illocutionary acts have the common purpose of generating a ‘normative state of affairs’ by determining the ‘rights, empowerments, permissions, obligations, and commitments of the participants in verbal interaction’.\textsuperscript{756} A successful illocution must have what Austin terms the necessary ‘felicity conditions’, such as the speaker intending to perform the act, doing so in the correct context and following the appropriate conventions.\textsuperscript{757} For example, to succeed as a declaration of marriage the speech act ‘I pronounce you husband and wife’ must be done in the appropriate context (a wedding), after going through the relevant conventions (a wedding ceremony) and is made by a speaker with authority (a priest).\textsuperscript{758} The absence of any of the felicity conditions means the illocution fails to generate a change in the normative state of affairs. Substituting a groomsman for the priest in the above example would mean the declaration of the marriage failed as an illocution, or what Austin would call a ‘misfire’.\textsuperscript{759} The pertinent felicity condition for the discussion that follows is the speaker’s authority. Authority here is descriptive, not normative, what Joseph Raz would term ‘effective’ or ‘de facto’ authority.\textsuperscript{760} The priest, the legislator, the manager, and the college president all have such authority by dint of rules that confer it on the positions they occupy. We can say that the legislator, manager, and the college president are abusing their positions, and their speech acts are unjust and illegitimate, but justness and legitimacy are normative criteria, not felicity conditions.\textsuperscript{761}

I will refer to this descriptive authority as practical authority.\textsuperscript{762} The legislator’s and manager’s practical authority is formal, but it can also be informal such as that of a church leader over their congregation or a community leader in a neighbourhood.\textsuperscript{763}

\textsuperscript{755} Langton (n 737) 305.
\textsuperscript{756} Maciej Witek, ‘How to Establish Authority with Words: Imperative Utterances and Presupposition Accommodation’ in Anna Brożek (ed), Logic, Methodology and Philosophy of Science at Warsaw University, Warszawa 2013 (2013) 155.
\textsuperscript{757} For an overview of felicity conditions including audience recognition see McGowan (n 288) 15–18.
\textsuperscript{758} Sanford Schane, ‘Contract Formation as a Speech Act’ in Lawrence M Solan and Peter M Tiersma (eds), The Oxford Handbook of Language and Law (Oxford University Press 2012) 101.
\textsuperscript{759}JL Austin, How to Do Things with Word (Harvard University Press 1975) 16.
\textsuperscript{761} Langton (n 742) 135–136. For an account of legitimate authority see Raz (n 760) ch 1.
\textsuperscript{762} I borrow this term from Langton (n 742) 125.
\textsuperscript{763} Maitra (n 752) 106.
Illocutions by one without authority, what Austin would term a ‘low type’, would misfire. A racist who is not a legislator in South Africa would not succeed in depriving Black South Africans of the right to vote, and someone other than the manager or college president would not succeed in creating these discriminatory hiring policies.

This has been referred to as the ‘authority problem’. Ordinary hate speakers (whom I will in this chapter refer to as ‘low types’) do not occupy any formal or informal position of authority, and their utterances lack a felicity condition necessary to be successful illocutions. Their speech acts cannot ordinarily constitute harm. The authority problem is one that is discussed extensively in the literature. I focus here on two contributions by Maitra and Langton who examine how low types can come to have authority. I rely primarily on Maitra’s framework and draw on Langton to bolster some weaknesses in her approach.

3. The limited authority of low types

Maitra conceives of two ways in which low types can come to possess the authority to engage in authoritative speech acts. The first is what she terms derived authority. This occurs when a person who possesses practical authority over the relevant domain confers it on a low type. She gives the example of a teacher who calls on a student to assign the other students tasks while the teacher is out of the classroom. The student’s speech acts ordering the other students to perform these tasks are successful illocutions by dint of this derived authority. The derived authority is conferred by an affirmative act by the teacher, but Maitra claims it can also be conferred by omission. She gives the example of a ‘bossy student’ who begins ordering students to perform tasks. The teacher is present but doesn’t intervene and allows the bossy student to carry on ordering the other students who then act on...
the bossy student’s instructions. The teacher’s omission (failure to intervene) confers authority on the bossy student.\footnote{ibid 105–106.}

Maitra claims that low types can also acquire derived authority through omissions. She gives the example of a black family that moves into a neighbourhood with known community leaders who have practical authority. The black family is then subjected to what Brown would term an ‘expression oriented-hate crime’ in the form of a burning cross anonymously left on their front lawn.\footnote{Brown, Hate Speech Law (n 55) 35.} Maitra notes that burning crosses have various symbolic meanings, but one of them is to rank the target as ‘undesirable’.\footnote{Maitra (n 752) 109.} The communities’ leaders then fail to denounce the act or express any sympathy towards the victims. This omission confers on the anonymous cross burner the derived authority to rank the victims. She acknowledges that these instances of derived authority would be scarce because, in most cases, it would be unclear who has practical authority over the relevant domain, whether they know about the incident at all and even if they did, what an appropriate denunciation would look like.\footnote{ibid 110.} Further, as Sumner points out, in liberal societies low types engaging in overt acts of racism, such as burning crosses, will often be denounced and challenged by community leaders and elected officials.\footnote{LW Sumner, ‘Reviewed Work(s): Speech and Harm: Controversies over Free Speech by Ishani Maitra and Mary Kate McGowan’ (2013) 39 Social Theory and Practice 710, 715.} It is only in exceptionally rare cases that low types will acquire such derived authority.

To account for a low types authority in other cases, Maitra argues that it is also possible for them to obtain it from the acts or omissions of those who lack practical authority through what she terms licensing.\footnote{Maitra (n 752) 107.} Licensing is a phenomenon whereby a low type acquires authority because other low types either accept their instructions or fail to challenge their authority. To explain how it occurs, Maitra and Langton rely on frameworks of linguistics pragmatics, specifically Robert Stalnaker’s common ground framework and David Lewis’s notion of a conversational score.\footnote{David Lewis, ‘Scorekeeping in a Language Game’ (1979) 8 Journal of Philosophical Logic 339; Robert Stalnaker, ‘Common Ground’ (2002) 25 Linguistics and Philosophy 701.} When parties engage in a conversation, they are pursuing the ‘common goal’ of successful
communication. In order to successfully communicate with one another, the conversational steps the speaker and hearer ought to take depends on what the other does; put simply, they must coordinate. If I ask, ‘what’s the weather like’ it is conversationally appropriate to respond with ‘it is warm’. It is not conversationally appropriate to respond with ‘I would also like some cheese’.

Presuppositions shared by the participants also determine what utterances are conversationally appropriate. This includes background information such as time of day, year, location etc. Maitra refers to these presuppositions known to all participants in the conversation as the ‘shared background’. If the shared background is that it is 4 pm, it would be conversationally inappropriate for me to ask, ‘what do you want for breakfast?’ However, the rules governing conversations are not rigid with the shared background changing and adapting as the conversation develops. Lewis refers to these as ‘rules of accommodation’. Utterances can introduce presuppositions which are accommodated and added to the shared background if they are not challenged or objected to by the other participants. For example, if I assert that ‘this is so easy even Ben could figure it out’, I have introduced a previously unshared presupposition that Ben is incompetent. Unless rejected or objected to, ‘what are you talking about? Ben is very smart’, it is accommodated and forms part of the shared background and ‘from then on it is appropriate for participants to continue with that presupposition in place’.

Maitra and Langton argue that authority can also be presupposed and accommodated into existence in a similar fashion. To demonstrate this, Maitra uses the example of a traffic jam caused by an accident where one driver (A) alights from their vehicle and begins to direct traffic, and the other drivers comply with their instructions. When A begins to direct traffic, they introduce the (previously unshared) presupposition that A has the authority to issue commands to the other drivers. By treating it as a binding order—and therefore not challenging the authority—this presupposition (that A has

---

777 McGowan (n 288) 11.
778 ibid.
779 Stalnaker (n 776) 701.
780 ‘Shared background’ is a term used by Maitra and I adopt that here to avoid back-and-forth references between this and Stalnaker’s term ‘common ground’ Maitra (n 752) 111.
781 Lewis (n 776).
782 Langton (n 742) 144.
783 Maitra (n 752) 106.
authority over the other drivers) is added to and becomes a part of the shared background. In other words, A’s authority ‘comes into being through accommodation’ or, as Witek would put it, A acquired the power to give the other drivers ‘felicitous directives’. Notice that all participants here are low types, and A only acquired practical authority after the other drivers failed to challenge it.

Maitra then demonstrates how this applies in a hate speech context. She sets out a scenario where an Arab woman in a subway car is accosted by an elderly white man who utters, ‘Fucking, terrorist, go home. We don’t need your kind here’. He then continues his verbal abuse while the woman and other passengers remain silent, despite them all being able to clearly hear his words. Because every passenger can hear the hate speaker, and they know that the other passengers can also hear him, they are all ‘(unwilling) participants in one conversation’. It is a conversation in which the speaker is seeking to introduce the presupposition that he has the authority to rank the woman as a ‘terrorist and an undesirable’ to the shared background. Maitra argues that because the participants in the conversation (all the other passengers in the subway car) fail to object to or reject the presupposition of authority, it is accommodated and becomes part of the shared background. This passivity grants the low type the authority required to rank the woman, and his hate speech constitutes her and their (Arabs) subordination.

Brown and Barendt refer to this solution to the authority problem as ‘ingenious’ but problematic in several respects, and similar criticisms have been levelled by Sumner. The first is that treating the failure to intervene as a form of implicit licensing places an unreasonable burden on the passengers who may rightly fear that any intervention on their part may prompt abuse or even physical violence from the hate speaker. Brown proposes the inclusion of a condition in Maitra’s framework where licensing can only be inferred where the audience has had a reasonable opportunity to reject/object to

784 Langton (n 742) 144.
785 Witek (n 756) 153.
786 Maitra (n 752) 115.
787 ibid 115.
788 ibid 115.
789 ibid.
790 Note that the other passengers don’t have to believe Arabs are inferior, merely to accept it for purposes of the conversation Maitra and McGowan (n 461) 116.
791 Brown, Hate Speech Law (n 55) 79; Sumner (n 774) 716.
the hate speaker and that such an opportunity cannot be deemed to exist where the audience has rational concerns about their safety.\textsuperscript{792}

Secondly, even if such licensing has occurred and the speaker does have authority to subordinate his target, there’s still a disanalogy between this authority and the practical authority possessed by the South African legislator, the company manager, or the college president. The legislator’s utterance is an exercise of Hohfeldian power as it removes a legal right (to vote) and therefore lowers the social status of Black South Africans as a matter of law.\textsuperscript{793} The manager and the college president have the power to subordinate Jews and women within their respective organisations.\textsuperscript{794} To have analogous authority, the low type’s speech act ‘Fucking terrorist, go home’ would need to remove rights or rank Arabs in some way, such as by prohibiting them from using the subway and the licensed authority Maitra describes does not give the hate speaker such authority.

Langton’s work on the authority problem can be used to bolster some of the weaknesses in Maitra’s account. She addresses this disanalogy between the legislator’s authoritative illocutions and that of the low type by relativising the notion of authority. She argues that all authority is relative to a subject domain, jurisdiction and to ‘rival authorities’.\textsuperscript{795} Take the manager’s utterance that ‘we aren’t going to be hiring any more Jews’. It’s an exercitive speech act ranking Jews as inferior and denying them employment. Its subject domain is limited to the treatment of Jews; it does not therefore also rank and exclude Black people, Asians, or Muslims from the organisation. Its jurisdiction is limited to a specific company, and it does not subordinate Jews in the wider society. It is very authoritative compared to rival authorities within the company by dint of the manager’s position in the organisational hierarchy. But in the wider society, this authority is subservient to any higher-order authorities, such as anti-discrimination laws that may be in force in the state.

\textsuperscript{792} Brown, Hate Speech Law (n 55) 79.  
\textsuperscript{793} Wesley Newcomb Hohfeld, ‘Fundamental Legal Conceptions as Applied in Judicial Reasoning’ (1917) 26 The Yale Law Journal 710.  
\textsuperscript{794} Sumner (n 774) 716.  
\textsuperscript{795} Langton (n 742) 125.
This demonstrates Langton’s point about relativity. The law has higher authority than other speech acts, and it ‘would be grotesque to suggest otherwise’.\footnote{ibid 137.} We can recognise that both the South African legislator and the manager have practical authority, consider the parallels between them without questioning the superiority of the former and acknowledge the ‘unique gravity’ of the law.\footnote{ibid.} In the same vein, we can recognise that low types can, in some cases, possess the authority to perform authoritative speech acts \emph{and} that this authority (whatever its scope) is inferior to that possessed by the legislator or the manager. Once the relativity of authority is conceded, we can account for instances of non-ideal authority, such as that potentially wielded by the low type on the subway car.\footnote{ibid.}

Langton’s other pertinent contribution is her more expansive conception of authority. She identifies another way of ‘authoritatively saying someone is inferior’ through the exercise of what she refers to as ‘epistemic authority’.\footnote{ibid 125.} The difference between practical and epistemic authority is explained by Joseph Raz who notes that practical authority is ‘authority for action’ whereas epistemic authority is ‘authority for belief’.\footnote{Raz (n 760) 8.} Practical authority is a ‘decision that something is to be so’ whereas epistemic authority is a ‘judgement that something is so’.\footnote{ibid 126.} The typical speech acts of practical authority are exercitives such as orders and directions, whereas the typical speech acts of epistemic authority are verdictives.\footnote{ibid 131.} Raz however argues that one can come to have practical authority by dint of their epistemic authority. An expert doctor both diagnoses an illness (epistemic authority) and prescribes a cure (practical authority).\footnote{ibid (n 760) 8.} While Raz emphasizes expertise as the source of epistemic authority, Langton claims that it can also be founded on credibility.\footnote{Langton (n 742) 141.} On Raz’s understanding, only the expert doctor has epistemic authority, whereas according to Langton, ‘the

\footnote{Langton (n 742) 141.}
\footnote{Raz refers to epistemic as ‘theoretical authority’ Raz (n 760) 8.}
\footnote{Langton (n 742) 126.}
\footnote{ibid 125.}
\footnote{Raz (n 760) 8.}
\footnote{Joseph Raz, The Authority of Law: Essays on Law and Morality (2 edition, Oxford University Press, USA 2009) 8; Langton (n 742) 131.}
credible quack doctor’ also has epistemic authority to diagnose and prescribe cures for their ‘gullible patients’. If credibility is a source of epistemic authority, we can say that the inverse to Raz’s formulation is possible; one can come to have epistemic authority from their practical authority. A finance minister’s practical authority gives them credibility on matters of finance. This practical authority gives them epistemic authority. Those with practical authority can also confer epistemic authority on others by endorsing their credibility. The finance minister could state, ‘Brian Smith’s forecasts for the stock market are sound’. Prior to this endorsement, Brian Smith was a mediocre forecaster. However, now we can say that he has credibility and, therefore, epistemic authority. This authority derives from the finance minister’s practical authority.

This analysis can now be expanded to include low types. Suppose that a person with practical authority officially endorses a low type, ‘that elderly man in the subway knows what he is talking about’, and the low type derives authority from this endorsement. Alternatively, the low type builds a cult following and becomes an expert to his gullible audience in much the same way as the ‘quack doctor’. In such cases, we can say that the low type is licensed by their audience. Either way, through either deriving it or being licensed, the low type acquires epistemic authority. With it, they can authoritatively convey the ‘fact’ that minority groups rank below the majority (epistemic authority) and direct that they should be treated accordingly (practical authority). These are not claims about the perlocutionary effect of the speech. It is not that the audience reads or listens to the hate speech, weighs the various pros and cons and is persuaded by the racial hierarchies the low type is selling. Rather, utterances by those with epistemic authority are trusted and provide strong reasons to update one’s beliefs to conform with them.

Linda Trinkaus Zagzebski refers to these as the two features of epistemic authority. The first feature is pre-emption which is ‘the fact that the authority has a belief \( p \) is a reason for me to believe \( p \) that replaces my other reasons relevant to believing \( p \) and

---

805 Langton (n 742) 142.
806 Zagzebski’s account of epistemic authority is also heavily influenced by Raz.
is not simply added to them’. The second is content independence where the reasons one has for adopting the authority's belief are not dependent on the content of the belief $p$. If instead the epistemic authority had believed $y$, one would have reason to believe $y$ instead of $p$. Zagzebski argues that where we have trust in an authority, i.e. find them credible, we believe what this authority tells us (pre-emptively and content-independently). Zagzebski uses the example of climate scientists, noting that we believe their judgements about the effects of greenhouse emissions, and if they had different beliefs about the effects of emissions, we would hold those instead. There are of course limits to this. As Adam Elga notes, one would not find credible a claim by a weather forecaster that it will rain eggplants tomorrow as this opinion falls ‘outside the appropriate’ range of opinions. Similarly, the most extreme claims made by hate speakers can also fall outside this range, even amongst those who find them credible; ‘Jews are climbing down your chimney every night and drinking your children’s blood’ whereas (comparatively) less absurd claims such as ‘Jews control the media’ may inform belief. From this, we can say that the low types who have epistemic authority can perform authoritative speech acts and that the audience who finds them credible believes them (pre-emptively and content-independently). This has deleterious effects on the dignity of the targeted because we know that social standing is determined both by law and informal social norms. It can be undermined formally by a legal racial caste system, such as apartheid legislation and informally through authoritative judgements that inform widespread beliefs about a group's inferiority.

With these contributions, we can reinforce Maitra’s framework to respond to some of her critics. Let us go back to her original example about the subway hate speaker and assume for the moment that he has been properly licensed by his audience and therefore has epistemic authority to rank the Arab woman. Using Langton’s three dimensions we can say that the speech act ‘Fucking, terrorist, go home. We don’t need

808 ibid.
809 ibid 109.
810 ibid 107.
813 Simpson (n 9) 716.
your kind here’ has a clear subject domain being the treatment of Arabs. Its jurisdiction is limited to the subway car, and because it has been licensed by those in the car, its authority compared to rivals is high.\textsuperscript{814} If these assumptions hold, we can say that the hate speaker has performed an exercitive speech act making the Arab woman (and perhaps Arabs more generally) rank below others within the subway car. This illocution does not prohibit Arabs from entering or staying in the subway car because ‘the authority of hate speech is ‘unlike the law in kind and degree’.\textsuperscript{815} Instead, the consequence of the audience’s failure to block/reject the speaker’s authority is that the target is left believing that the other passengers agree with the hate speaker’s ranking. The speech act, therefore, makes the subway car a hostile environment for the target, and it, therefore, constitutes their subordination.\textsuperscript{816}

If, however—as is more likely the case—Brown and Sumner are correct and mere silence cannot be sufficient for licensing, the hate speaker may still possess a narrower epistemic authority through credibility. The subject domain would remain the same, the jurisdiction is limited only to those in the subway car who find the speaker credible; likely only amongst high-prejudice individuals and its authority amongst rivals may be strong among this subset of people. While a crude hate speaker may well be a ‘detested lone wolf’ amongst the egalitarian minded, they may still occupy an esteemed status and high credibility amongst their fellow haters, as Langton notes:

\textit{a low type may be all one has when it comes to authority. ‘In the land of the blind, the one-eyed man is king’, as the unreconstructed proverb has it. If hate speakers are the only local voices, whether because of state-imposed restrictions, or technology-imposed echo-chambers, then hate speakers will have the authority of a monopoly.}\textsuperscript{817}

We can say that the low type’s speech act is a verdictive, authoritatively conveying the ‘fact’ that Arabs are inferior amongst the subset of those in the subway car who regard him as credible. Notice how this has narrowed the scope of the authority significantly. In fact, it may be the case that there are no fellow bigots in the subway car at all, in which case the speaker would not have any authority, and their speech act would not

\textsuperscript{814} Langton (n 742) 138–139.
\textsuperscript{815} McGowan (n 288) 174; Langton (n 742) 147.
\textsuperscript{816} Maitra (n 752) 117.
\textsuperscript{817} Langton (n 742) 140; Waldron, \textit{The Harm in Hate Speech} (n 14) 198.
have the felicity conditions required for a successful illocution. This means that any harm inflicted, such as psychological harm or undermining dignity, could only be caused by the speech, *not* constituted by it. Therefore, save for cases where hate speakers have practical authority from occupying a position of authority, derived authority from endorsement by those with practical authority, clearly licensed authority through affirmative actions by others, or are an influential hate ‘guru’ with a large following, low types will rarely have the requisite authority to engage in authoritative speech acts.  

818 Does this hold true for the press?

4. The press as authoritative speakers

Ascertaining this requires answering two questions. Firstly, can covert hate speech such as DNS subordinate? And secondly, what is the nature and scope (if any) of the press’s authority to engage in authoritative speech acts?

Andrew Altman claims that there is a distinction between overt and covert hate speech in terms of the capacity to subordinate. He writes, ‘the primary verbal instruments for treating people as moral subordinates are the slurs and epithets of hate speech’.  

819 He contrasts this with covert hate speech which he describes as the ‘scientific’ or ‘philosophical’ kind in which category he includes arguments of innate genetic or cultural inferiority between groups.  

820 Altman argues that calling someone a ‘faggot’ is a more powerful speech act than stating ‘you are contemptible for being homosexual’, even if both constitute subordination.  

821 Here Altman seems to conflate causal and constituted harms. We can grant that overt hate speech is worse in some respects. We have seen it is more likely to *cause* psychological harm and severe-emotional distress. However, this has little to do with whether it constitutes harm. When someone exercises practical authority to rank groups as lower or epistemic authority to authoritatively judge them as inferior, surely the problem is in the ranking *qua* ranking rather than whether overt expressions are used. One manager enacts a new hiring policy by stating, ‘No more Niggers, they are too thick to cut it’. Another enacts the same policy by asserting ‘*, I would hire Blacks, but this role requires a certain kind of

---

818 Langton (n 742) 146–148.
819 Altman (n 750) 311.
820 ibid.
821 ibid.
person, possessing certain skills and I don’t think Blacks are ready for that just now’. Both speech acts constitute subordination regardless of the pseudo-civil nature of the latter.

This explains why Altman’s position is an outlier. Brown cites both overt hate speech (group defamation and epithets) as well as negative stereotyping as categories of speech that can be used to subordinate. Similarly, Langton argues that negative stereotyping (what she terms ‘generics’) appear ‘innocent but may nevertheless undergird racial subordination’. She then gives examples of what I would categorise as DNS, such as ‘Latinos are Lazy, Muslims are violent’ and ‘Blacks are sexually insatiable’. The answer to the first question then is in the affirmative; covert hate speech, including negative stereotyping by the press, can subordinate.

Turning now to the second question, what is the nature and scope of the press’s authority? One plausible view is that, unlike low types, the press has derived authority. Recall that one can acquire authority when it is conferred on them through commission or omission by someone with practical authority. We have seen that endorsement by someone with practical authority (the finance minister) can be the source of another’s epistemic authority (the mediocre forecaster).

Langton illustrates how such endorsement can confer authority on publications using the example of Der Stürmer, a virulently antisemitic newspaper published between 1923-1945 by Julius Streicher, a member of the Nazi party convicted of crimes against humanity at the Nuremberg tribunals. It was not an official publication of the party, with even prominent members split on it, but it had the endorsement of its highest echelons. Using Langton’s three dimensions, we can say that its subject matter was about a group of people (Jews), and it had a specific jurisdiction (its readers). Its credibility amongst rivals was high partly because of a lack of credible alternatives in

---

822 Brown, Hate Speech Law (n 55) 80.
823 Rae Langton, Sally Haslanger and Luvell Anderson, ‘Language and Race’ in Gillian Russell and Delia Graff Fara (eds), The Routledge Companion to Philosophy of Language (Routledge 2013) 760.
825 Hermann Wilhelm Göring and Joseph Goebbels were amongst the high-ranking Nazi party figures who disliked the publication, but it was supported by Adolf Hitler. For an overview of the rise and fall of Der Stürmer and Streicher see Randall Lee Bytwerk, Julius Streicher (1976); Christian Zentner and Friedemann Bedurftig, The Encyclopedia of the Third Reich (Da Capo Press Inc 1997).
Nazi Germany and partly because of endorsement from those with practical authority which Langton argues ‘enabled it to masquerade as testimony, conveying expert knowledge of a Jewish menace’. This endorsement is what gave the publication epistemic authority and enabled it to authoritatively convey the ‘fact’ that Jews are inferior. Because epistemic authority is also a source of practical authority, it also allowed the publication to direct its readers to treat Jews as inferior. Therefore, it constituted the subordination of Jews.

This can be applied to the press. This is not to suggest that the contents of Der Stürmer and that produced by the British press are comparable, but rather that by applying Langton’s analysis, we can see how endorsement can provide publications with derived authority to engage in authoritative acts of hate speech. We have seen that low types will generally not have this type of authority because, in liberal societies, their speech acts are rejected and condemned rather than endorsed by those with practical authority. However, the same does not apply to the press who are often endorsed by those with practical authority. This endorsement is often more implicit than the finance minister’s endorsement of the mediocre forecaster.

An example of implicit endorsement is where those with practical authority associate themselves with a publication by writing for them. There are various examples of high-ranking government officials writing editorials for various press outlets. In doing so, they are not just communicating their views on various political issues to the outlet’s readers. They are also implicitly endorsing its credibility. If a minister in charge of immigration writes for an outlet that has an anti-immigration editorial policy, they are lending legitimacy, respectability, and epistemic authority to that outlet. The same thing occurs at a regional or even local level when say, a member of parliament or other leader writes an article for a local outlet to communicate with their constituents.

---

826 Langton (n 742) 138.
Another form of implicit endorsement is through those with practical authority actively courting and seeking support from a press outlet. The claim that seeking endorsement is itself an endorsement of the endorsee may seem counterintuitive on its face, but by seeking endorsement, one is implicitly claiming that a publication is credible, legitimate, and capable of shaping public opinion. Curran and Seaton detail the extent to which various British political parties have, over the years, sought the endorsement of the institutional press in electoral campaigns and to build support for their policies.\textsuperscript{828} By treating the press as kingmakers, those with practical authority have made them kingmakers and given them the credibility necessary to engage in authoritative speech acts.

Lastly, advertising can also be a form of implicit endorsement. The purpose of advertising is to promote a business and its products and services. Therefore, by choosing to run advertisements on a particular platform, a business is claiming that the platform would succeed in promoting its brand, and it is therefore an acknowledgement that this platform is credible and legitimate. Conversely, businesses avoid advertising on platforms that they deem illegitimate as they fear that associating with them would bring their brand into disrepute.\textsuperscript{829} For example, several businesses suspended their advertisements from \textit{GB News TV}, a British television and radio news channel, alleging that the controversial views expressed on the channel contradicted their values.\textsuperscript{830} However, advertising is not limited to businesses as governments regularly use advertisements to bring attention to various policies and campaigns or to widely distribute messages to the public. The British government has run several such advertisements on both traditional and new media, including the press.\textsuperscript{831} It is therefore arguable that the government’s choice to advertise on these outlets amounts to an implicit endorsement of their credibility. One response to this is that, unlike

\begin{footnotes}
\item Curran and Seaton (n 26) ch 9.
\end{footnotes}
businesses, governments are not selling wares or building brands. Rather, their goal is to convey information to as many people as possible, and this means the principal criteria for selecting outlets is their reach, not their values. However, even if this is the case, government procurement is regulated, and the criterion for selecting suppliers includes an assessment of whether they operate legally, ethically, and promote diversity and inclusion. By advertising on these outlets, the government is claiming that these outlets align with these values, and this amounts to an endorsement of their credibility.

If I am correct, then we can say that some press outlets have epistemic authority which is derived from endorsement by those with practical authority. Here we must take a nuanced and relativised understanding of authority. These outlets do not derive power to make and enforce laws from government endorsement. Rather, they derive credibility, which as we have seen, can be a source of epistemic and practical authority.

Let us now see how this applies to the subway car example. Assume that the statement ‘Fucking, terrorists, go home. We don’t need your kind here’ appeared in the pages of a publication endorsed by someone with practical authority, such as a government minister. Applying Langton’s dimensions, we can say that the subject matter is about a group of people (Arabs), it has a specific jurisdiction (its readers), and its credibility amongst rival publications is high because of its endorsement by those with practical authority. This allows the paper to authoritatively tell its readers the ‘fact’ that Arabs are inferior, and this informs their beliefs pre-emptively and content-independently. The speech act constitutes the subordination of Arabs.

Another plausible view is that the press’s authority is licensed by their readers. Recall that the primary weakness in Maitra’s claim was that she treated the silence of the passengers in the subway car as a form of licensing of the elderly man’s racial abuse. If Brown is correct and licensing requires some affirmative act, it can be argued that buying physical copies of newspapers, paying for digital subscriptions or even visiting news websites are all affirmative acts that support a particular outlet and can therefore

---

amount to licensing their speech. In fact, the press often claims their right to be partisan on the basis that they represent and speak on behalf of their audience. At the very least, these acts by readers are far more affirmative than those of the passive subway passengers. This licensing gives an outlet the authority to rank its target. Once again, applying Langton’s three dimensions to the subway example, we can say that the subject matter and jurisdiction are the same, and its credibility amongst rivals is high because it has been licensed by its readers. The consequence of this licensing is that the targeted group (Arabs) come to believe that the publication’s speech act reflects the views of its readers. Therefore, it creates a hostile environment for Arabs amongst its readership and constitutes their subordination.

Finally, similar to low types, the press also has epistemic authority amongst those who find them credible. Recent surveys by Ofcom indicate that 65% of regular newspaper consumers find them trustworthy. We have seen that newspapers (including print, website and app versions) are the principal source of news for 47% of adults in the UK. Recall also that three companies, namely DMG Media, Reach and News UK, account for 90% of readership figures.

I will now apply Langton’s dimensions to the subway example to see the effect of the press’s epistemic authority. The subject matter is unchanged, the jurisdiction is narrower, limited only to the subset of the readership who find the press to be trustworthy and therefore credible; authority amongst rivals may be strong partly because this subset finds them credible and partly because of the absence of rival authorities in a concentrated newspaper market. This operates on different scales. For example, local newspapers may not have a wide circulation, but they may have significant credibility in the local community, similar to the community leaders in Maitra’s example earlier. Whereas a low type also has epistemic authority, this is limited to their fellow fringe extremists who find them credible. On the other hand, the national, regional, and in some cases, the local press have credibility amongst a far

---

833 Newspapers generate advertising revenue from pay-per-click models. For an overview of these different funding models see Peter S Cohan, ‘Newspapers’ in Peter S Cohan (ed), Goliath Strikes Back: How Traditional Retailers Are Winning Back Customers from Ecommerce Startups (Apress 2020) 51–69.
834 Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 616.
835 ‘News Consumption in the UK: 2022’ (n 50) 72.
836 ibid 13.
837 Chivers (n 375) 2.
greater number of people. As such, their authoritative judgements are more consequential for a group's social standing nationally, regionally, and locally.

This comparison demonstrates that in most cases, low types will not have the authority necessary to successfully rank and subordinate their targets. They do not occupy positions giving them practical authority. They also do not have derived authority, as those who occupy such positions will not usually confer such authority on them. Their speech acts will rarely be licensed both because silence does not amount to licensing and because their speech acts are widely challenged and therefore blocked. Finally, any epistemic authority they have is limited and only exists among fellow low types who find them credible. Their speech acts rarely constitute harm. On the other hand, we have seen that some press outlets have derived authority from endorsement by those with practical authority, and all have licensed authority through support from their readership. They also have much broader epistemic authority as they are regarded as credible by a significant number of adults in the UK. Their speech acts often constitute harm.

Finally, the fact that the press has this kind of authority has implications on their choices to uncritically platform low types. Providing column inches to fringe extremists could reasonably be understood as an endorsement or licensing of a low type. This lends the extremist credibility and thus the authority necessary for their hate speech to constitute harm. This argument is not confined to fringe extremists as the epistemic authority of mainstream figures can also be broadened through such platforming.

5. The consequences of constituting harm: enacting harmful social norms

Having established that hate speech by the press also constitutes harm, I will now consider the consequences that flow from this. One view is that from the target's perspective, there is no real difference between hate speech causing and constituting harm. To them, the pertinent issue is that their social standing has been lowered, and it is largely immaterial whether this is achieved through illocutions or perlocutions. Rather, the real difference is that speech that constitutes harm ought not to be covered by a free speech principle. Kent Greenawalt would term most illocutionary acts as
‘situation altering utterances’.\textsuperscript{838} Marriage vows, contracts, promises, and an umpire’s shouting ‘out’ either create legal obligations, alter a person’s normative obligations, or change the social context.\textsuperscript{839} He argues that ‘such utterances are ways of doing things, not of asserting things’.\textsuperscript{840} Therefore, they do not count as speech or expression for purposes of a free speech principle as they are unconnected to the values underpinning the principle. As Barendt puts it:

\begin{quote}
the reasons for treating speech as special under the free speech principle simply do not apply to situation-altering utterances; a contractual promise or bet does not contribute to the discovery of truth or to the working of a participatory democracy.\textsuperscript{841}
\end{quote}

This is predicated on Schauer’s coverage/protection distinction which was discussed in an earlier chapter, and I will not rehash it here. The claim here is simply that situation-altering utterances are forms of conduct rather than speech and ought to fall outside the coverage of a free speech principle. Recall that, unlike conduct, speech enjoys special immunity from regulation under a free speech principle.\textsuperscript{842} Since hate speech constituting harm is conduct, it is also regulable on a ‘lesser showing of harm’, including on purely aesthetic grounds and offensiveness.\textsuperscript{843} Barendt notes that such an approach means lawmakers are not required to engage in any principled balancing between the values of the (uncovered) speech against the harm caused.\textsuperscript{844} This is the approach taken by Langton, who argues that authoritative acts of hate speech ‘are not the sort of speech a free speech principle protects’.\textsuperscript{845}

The implication of this on press hate speech is significant. Since press hate speech often constitutes harm, and speech constituting harm is uncovered by a free speech principle, it amounts to no-value speech, and there is significant room for its curtailment by legislators. It would, in principle, support imposing prohibitions on

\begin{footnotes}
\textsuperscript{838} Greenawalt (n 287) 57–58.
\textsuperscript{839} ibid 57–58.
\textsuperscript{840} ibid 58.
\textsuperscript{841} Barendt, ‘What Is the Harm of Hate Speech?’ (n 289) 550–551.
\textsuperscript{842} Greenawalt (n 287) 58.
\textsuperscript{843} McGowan (n 288) 163.
\textsuperscript{844} Barendt, ‘What Is the Harm of Hate Speech?’ (n 289) 551.
\textsuperscript{845} Langton (n 742) 129.
\end{footnotes}
offensive language and the broader hate speech provisions currently applicable only
to broadcasters to the press.846

However, this approach is problematic. Firstly, Greenawalt himself clearly believed
that hate speech ought to be covered by a free speech principle, a view shared by
Barendt.847 He argues that hate speech discloses the speaker's beliefs and values,
however pernicious those values may be.848 Secondly, all utterances are situation-
altering to some degree, and this does not automatically deprive them of coverage.
Greenawalt notes that even in cases of non-authoritative speech, ‘the world has
changed to the degree that the speaker has now stated something he had not
previously stated in the precise context’.849 He qualifies his claim by arguing that only
utterances ‘which are dominantly and clearly situation-altering’ ought not to be covered
by a free speech principle.850 If I enter into a contract with you, I may be disclosing my
belief that you are someone I want to do business with, but a contract is primarily a
situation-altering utterance that creates legal obligations.

Similarly, the legislator’s utterance depriving Black South Africans of the right to vote
and the manager’s policy enacting discriminatory hiring practices both disclose their
beliefs that these groups are inferior, but their primary function is situation-altering by
enacting changes to the legal order and companywide hiring practices respectively.
All of these are rightly uncovered under a free speech principle. On the other hand,
hate speech such as ‘Fucking, terrorist, go home. We don’t need your kind here’,
whether stated by an authoritative speaker or not, primarily discloses the speaker’s
values and therefore ought to be covered by a free speech principle.

A more promising view of the effect of constituted harms is offered by McGowan.
Firstly, she defines it as follows: ‘Harm constitution is a technical notion; it is a distinct
way of causing harm. Harm is constituted when the harm is brought about via
adherence to norms enacted’.851 To constitute harm, illocutionary acts must enact
harmful norms (what she terms ‘permissibility facts’), the norm must be followed and

---

846 Ofcom Broadcasting Code, s. (1), (2).
848 Greenawalt (n 287) 300–301.
849 ibid 61.
850 ibid 63.
851 McGowan (n 288) 2.
following it must result in harm. On McGowan’s understanding, the harm is still ‘causally downstream’ from the utterance. I will apply this to the examples of practical and epistemic authority discussed thus far and set out its implications.

The South African legislator’s utterance enacts a legal norm. It enacts strong permissibility facts that must be followed. Following this legal norm results in the disenfranchisement of South Africans. Electoral officials who fail to follow this legal norm by allowing Black South Africans to vote may suffer penal consequences. The manager’s utterance enacts norms governing hiring policies, and these norms also enact strong permissibility facts; organisational rules which must be followed. Following these norms results in the harm of discrimination. Employees who fail to follow these discriminatory policies in the hiring process may suffer disciplinary consequences. The primary purpose of these utterances is to enact legal norms and hiring policies respectively. Therefore, they are dominantly situation-altering and fall outside the coverage of a free speech principle.

On the other hand, press hate speech enacts social norms. McGowan describes social norms as:

*prescriptive: they give rise to talk about what a person ought to do; they afford incentives and reasons for action; they are used to explain behavior, and they ground criticism of those who do not comply.*

I have already argued that the authority of hate speech is not comparable to the authority of law and hiring policies in degree. Similarly, social norms enact weaker permissibility facts than legal norms and hiring policies. McGowan also concedes that, unlike these other examples, the primary function of hate speech is to disclose the speaker’s values and is therefore covered by a free speech principle. Instead, she argues that it is ‘harmful enough to warrant legal intervention’ even under a free speech principle.

---

852 ibid 24–25.
853 ibid 24.
854 ibid 25.
855 ibid 1.
With these differences in mind, I will now look at an example of how hate speech enacts social norms. Suppose a press outlet with relevant epistemic authority publishes the statement that ‘Muslims are culturally backward and prone to violence. We should monitor them’. It has authoritatively conveyed the ‘fact’ that Muslims are inferior and violent and tells its readers what to do about it. It enacts harmful social norms. Following these social norms (believing and treating Muslims as inferior) constitutes the harm of subordination. Failure to follow these norms results in ‘social censure’ for non-compliance.\(^\text{856}\)

Two things stand out from this concept of constituting harm. Firstly, once enacted, these social norms can be pernicious because they can either be resistant to/or disincentivise counter-speech. They are resistant to the victim’s counter-speech because the speech act has already diminished their social standing, and we have already seen that the speech of the subordinated receives less consideration by the majority.\(^\text{857}\) Counter-speech by members of the majority group may be effective, but the social norms enacted both pressures them to comply and often disincentivises such speech. Social norms have less prescriptive force than either legal norms or a hiring policy because these are backed by the enforcement powers of the state and the organisation respectively. However, they are still powerful and are routinely followed. Research examining user-generated comment sections of online news outlets finds that reader comments follow the norms set by the framing of the story. Where an article presents certain groups as dangerous or inferior, user comments will also be hostile towards them.\(^\text{858}\) Langton writes about the hate speech directed towards Gina Miller, a Guyanese-British activist who led a successful legal challenge to the UK government’s unlawful prorogation of parliament. As a result of this, Miller faced racist abuse and vile rape and death threats by online mobs. She recounts how the initial barrage of abuse normalised hatred against her and emboldened hate speakers to engage in even more extreme forms of abuse.\(^\text{859}\) Langton writes about this in the context of how the speech act of low types can—through accommodation—normalise hatred towards others. What Langton fails to note though is that there was

---

\(^{856}\) ibid 26.
\(^{857}\) Brown, *Hate Speech Law* (n 55) 84–85; Matsuda (n 510) 69.
\(^{859}\) Langton (n 737) 128.
relatively little online discussion about Miller until Brexit-supporting newspapers ‘went to town’ on her, leading to a major spike in hate speech directed at her.\textsuperscript{860} It was press coverage referring to her as an ‘ageing trollop’ and a ‘snake-in-the-grass’ which painted a target on her back by enacting norms that made her a legitimate object of abuse for their readers who dutifully obliged.\textsuperscript{861}

Further, social norms are stifling as even those readers with different views will often not air them for fear of social sanction from fellow readers. McGowan argues that ‘individuals routinely go against self-interest to avoid the social censure’ that results from a failure to comply with social norms.\textsuperscript{862} Lawrence Lessig gives an example of the social norm of pistol duelling by men in the American South in the 17\textsuperscript{th} and 18\textsuperscript{th} centuries. Men would risk death at the most trivial of insults.\textsuperscript{863} Many were willing to accept this risk simply because the refusal to accede to a challenge ‘opened one up to severe social sanction’.\textsuperscript{864} Similar consequences flow from authoritative acts of hate speech. Rating systems in comments sections, where others can ‘downvote’ comments are particularly effective at sanctioning dissent.\textsuperscript{865} Alternative viewpoints can attract denunciations and invite a pile-on from other readers.\textsuperscript{866} This makes constituted harms more problematic as unlike speech which causes harm, ‘they can be hard to answer with more speech’.\textsuperscript{867}

Secondly, unlike these direct and indirect harms explored earlier, causation is easier to establish. We have seen that many of the direct and indirect harms caused by hate speech involved complicated chains of causation with many intervening variables that make it difficult to assign and measure the contribution played by hate speech. On the other hand, constituted harms are more straightforward. As Brown notes, they are ‘defined in terms of sets of formal linguistic qualities occurring under sets of

\textsuperscript{860} John Mair and Mark Thompson, \textit{Anti-Social Media?: The Impact on Journalism and Society} (Tor Clark and others eds, Abramis 2020) 10.
\textsuperscript{861} ibid 13.
\textsuperscript{862} McGowan (n 288) 26.
\textsuperscript{863} Lawrence gives the example of a person behaving ‘coldly’ towards another as justification for a duel Lawrence Lessig, ‘The Regulation of Social Meaning’ (1995) 62 The University of Chicago Law Review 943, 969.
\textsuperscript{864} ibid 970.
\textsuperscript{866} ibid.
\textsuperscript{867} Langton (n 742) 147.
sociolinguistic conditions’ which are ‘amenable to social scientific assessment by linguists and sociologists’. The criteria or ‘felicity conditions’ necessary for successful illocutionary acts are known. To determine whether hate speech has enacted harmful social norms, we simply ask whether it has met these formal conditions. We have seen that the press often meets them, and low types rarely do. There are certainly some areas of uncertainty, particularly in the relative nature of authority and norms enacted. However, even taking these into account, we are still able to make more definitive claims about the harmful consequences of hate speech under the frameworks of linguistic pragmatics than we are under the traditional causal methods examined earlier.

6. Conclusion

To justify broader restrictions under the equivalence model, it was necessary to demonstrate that DNS by the press causes greater harms than overt hate speech by ordinary hate speakers. In chapters 4 and 5, a review of the empirical evidence suggests that this is indeed the cause. We have seen that hate speech by the press causes these various direct and indirect harms to a greater degree than that of the ordinary hate speaker. The case for regulating this speech has, therefore, already been made out.

This chapter has gone further, demonstrating how the press is unique in that unlike an ordinary hate speaker, their speech acts can also constitute harm. The ordinary hate speaker is a low type. Their targeted hate speech may directly harm a victim. Their diffuse hate speech may persuade an audience to adopt hostile attitudes towards vulnerable minorities and contribute to the creation of a climate of hatred. However, they do not have the authority required for their speech acts to authoritatively rank and subordinate their targets.

Conversely, DNS by the press often constitutes harm. They have the requisite authority to engage in authoritative acts of hate speech. This authority is derived, licensed and epistemic. Their speech acts enact harmful social norms, providing

---

authority for belief by telling us that certain groups are inferior and authority for action directing us to treat them as inferior.

These social norms are more resistant to strategies that may be effective against causal harms such as counter-speech, and the fact of their enactment is easier to ascertain than often indirect and complex causal harms. The harms of press hate speech are multidimensional, broader and more pernicious.

Broader regulation of press hate speech is therefore entirely compatible with, nay, required by the equivalence model. Current line-drawing unduly privileges press hate speech and therefore amounts to special treatment. Equivalence necessitates broader regulation of press hate speech. To be treated the same, the press must be treated differently. The challenges posed to broader hate speech regulation by the libertarian theory and its equivalence model have been overcome.
Chapter 7: Social Responsibility Theories and a Free Speech Principle

1. Introduction

In chapters 4, 5 and 6, I examined and set out the various direct, indirect, and constituted harms caused by DNS published by the press and how their regulation is compatible with the libertarian theory of press freedom and its equivalence model. In the next two chapters, I consider whether such regulation is also supported by SRT, specifically the variable geometry model of press freedom set out by Fenwick and Phillipson.

Section 2 briefly recaps the conditions under which broader regulation of press hate speech is compatible with SRT. Sections 3, 4 and 5 examine the various justifications for treating hate speech as high-value speech under a free speech principle. It will be argued that the most persuasive arguments for doing so are non-instrumentalist speaker-based justifications which either do not apply to or are weaker in the case of the press. Under this model, DNS by the press amounts to low-value or no-value speech which can be subject to broader restrictions than those imposed on hate speech by lone hate speakers.

2. The SRT Challenge to hate speech regulation

We have seen that SRT treat individual freedom of expression and press freedom as distinct principles. I have also argued that under SRT, coercive restrictions on hate speech that go beyond those imposed on the public can be applied to the press if both of the following conditions pertain: 869

i. The press ought to have more limited rights to engage in hate speech than lone hate speakers; and

ii. Such restrictions do not hinder the press’s ability to perform their democratic functions.

I deal with the first condition in this chapter and the second in chapter 8.

869 These conditions are explained earlier in Chapter 2
a. Optimal balancing under the variable geometry model

Recall that speech at the core of a free speech principle ought to be assigned greater weight in the balancing exercise, and I refer to this as high-value speech. Recall also that by weight, I mean degree of importance and balancing is simply moral reasoning conducted by legislators to determine the appropriate scope of a statute and/or by the courts in determining which values ought to take priority over the other in the circumstances of a particular case. High-value speech creates a presumption in favour of protection. As Brison argues, it justifies ‘weighing the interests with a thumb on the scale in favour of free speech’. It means that any justifiable restrictions on high-value speech—such as preventing harm to others—must be narrowly drawn.

There’s a consensus amongst political theorists and in media law scholarship that the most persuasive arguments for treating hate speech as high-value speech are non-instrumentalist speaker-based arguments. Waldron evaluates these arguments, specifically Baker’s argument from autonomous self-disclosure as well as Dworkin’s and Weinstein’s arguments from a citizen’s right to political participation, setting out the implications they have on the scope of hate speech restrictions. It will be argued that under these accounts, hate speech by members of the public ought to be treated as high-value speech. Waldron argues that in such cases, restrictions on hate speech must be narrowly drawn, limited only to overt hate speech and cites the UK’s stirring up offences as an example of legislation that meets these requirements. To Waldron, such legislation strikes an optimal balance between the rights of the hate speaker and the prevention of harm. This is a view I endorse and is the assumption that underpins my analysis.

In chapters 4, 5 and 6, we have seen that broader restrictions can justifiably be imposed on the press even if we treat their hate speech as high-value speech under a free speech principle. This is because of the press’s unique capacity to cause and constitute greater harm than lone hate speakers. In some sense, condition (i) above

---

870 Brison (n 346) 319.
872 Waldron, The Harm in Hate Speech (n 14) 8, 9, 149.
873 Barendt expresses a similar view Barendt, Freedom of Speech (n 6) 38.
874 See discussion in chapter 2
is superfluous. However, it is still important to evaluate the extent to which free speech theories support press claims to a right to engage in hate speech. This allows for an exploration of the motivations informing the publication of such content, which in turn inform the regulatory proposals developed later.\textsuperscript{875} It also serves as a supplementary argument for broader regulation should SRT reject the claim that press hate speech produces greater harm than that of lone hate speakers. Lastly, this is a matter which is underexplored in the broader hate speech literature and is worthy of examination in its own right.

It will be argued that hate speech by the press is not high-value speech that deserves strong protection under a free speech principle. I will demonstrate that the non-instrumentalist speaker-based justifications for treating hate speech as high-value speech are either inapplicable to, or are weaker in the cases of the institutional press. If this is the case, then press hate speech is, at best, either at the margins of a free speech principle (low-value speech) or is entirely uncovered (no-value speech). Recall that no-value speech does not benefit from any special immunity from regulation and is subject to the same justification for regulation as any other domain of human activity. Low-value speech, on the other hand, benefits from some protection, but it does not have the same weight as high-value speech and restrictions against it need not be so narrowly drawn. Net-interest realisation or optimal balancing in such cases requires moving up the scale of constraints, i.e. instituting broader bans than the stirring up offences.

3. Non-instrumentalist speaker-based accounts for free speech

I will now consider two non-instrumentalist accounts for a free speech principle: Baker’s argument from self-disclosure as well as Dworkin’s and Weinstein’s arguments from a citizen’s right to political participation, arguing that both are powerful accounts which, if applicable, make it difficult to restrict DNS without destroying political legitimacy. I will suggest that these arguments are inapplicable to—or are weaker when applied to—the press. As such, broader regulation of press hate speech does not have a similar deleterious impact on political legitimacy. Finally, I will consider

\textsuperscript{875} See chapter 9
whether the press’s claims to engage in hate speech can instead be justified under Scanlon’s listener-based autonomy argument as well as the argument from self-fulfilment.

a. Baker’s argument from self-disclosure: A right to be hateful

Baker argues that the leeway granted to individuals to engage in hate speech impacts on political legitimacy. Political legitimacy here is used in the normative sense to refer to the criteria that generate moral obligations on citizens to obey the law and which justify the state coercively enforcing such laws against them. These criteria are contested. Baker suggests one of them is that the state must respect the autonomy of ‘the people whom it asks to obey its laws’. Autonomy is a vague term, and Baker attempts to confine it, arguing that it must mean something more than allowing individuals to do whatever they please as it then becomes indistinguishable from general libertarian claims for freedom of action and choice and makes autonomy ‘virtually meaningless’ as a justification for a free speech principle.

Baker separates autonomy into two types, only one of which has relevance to political legitimacy. Firstly is substantive autonomy, which refers to a person’s ‘capacity (including the necessary opportunities) to lead a meaningfully self-authored life without unnecessary or inappropriate frustration by others’. This autonomy is never perfectly realised and exists on a continuum because a person’s capacity to lead a self-authored life requires ‘material resources, psychological resources, and other natural and social conditions’. Baker argues that education, access to information, favourable environmental policies and redistribution of wealth all enhance substantive

---


879 ibid 252.


autonomy. All of these entail the allocation of resources, and the relevant political organs need only satisfy the baseline-standard justification to pursue them.

Baker argues that the enhancement of everyone’s substantive autonomy ought to be a primary goal pursued by the state, but it must be done in a way that respects formal autonomy. This he defines as ‘a person’s authority (or right) to make decisions about herself—her own meaningful actions and usually her use of her resources—if her actions do not block others’ similar authority or rights.’ Unlike substantive autonomy, respect by the state for an individual’s formal autonomy is an ‘on/off’ valve. It either does, or it does not.

Respecting formal autonomy means allocating control over an individual’s mind only to themselves. It means allowing them to decide for themselves what to say, when to say it and whom to say it to. This, Baker argues, includes a ‘right to seek to persuade or unite or associate with others—or to offend, expose, condemn, or disassociate with them’.

What is compelling about Baker’s argument is that it allocates the value of the speech to the speaker as a manifestation and representation of who they are to the rest of the world rather than the value such speech holds to the rest of society. This is especially relevant to hate speech which is often low-value speech to most people other than the speaker. The notion that some people are worth less than others, that they are terrorists or apes, are not valuable contributions to public discourse. Rather the strongest case that can be made for protecting such speech is that it is valued in spite of rather than because of its contribution to public discourse.

A racist places great value in their distaste of minorities; by engaging in hate speech, they ‘want to present’ themselves and their personalities to others. This disclosure of their values, its viciousness and manner of delivery is what the speaker is serving

883 The notion of a baseline standard is explored in chapter 2.
886 ibid 225.
888 Waldron, The Harm in Hate Speech (n 14) 157, 164.
up to the world for others to respond to. It is a naked embodiment of their values ‘without distraction’.\textsuperscript{889} We recognise that speech in this way is centrally connected to the racist’s formal autonomy, and therefore enacting laws disallowing them from disclosing their values on their own terms infringes their formal autonomy.\textsuperscript{890}

However powerful Baker’s account may be, our concerns about the harms brought about by allowing such speech to circulate unregulated do not suddenly disappear. How then does Baker’s account fit into a balancing approach? At first, these may seem to be irreconcilable positions. Protecting a racist’s right to espouse their racism would cause and constitute the various harms set out in earlier chapters and prohibiting them from doing so would violate their formal autonomy and erode political legitimacy. I will later argue that legitimacy exists on a spectrum and is consistent with balancing values. However, for the moment, assume that legislators are faced with the task of balancing the speaker’s formal autonomy and the need to prevent the harms inflicted by hate speech. Waldron argues that the ideal compromise is reflected in legislation that bans only overt hate speech. Such legislation allows the speaker to ‘let off steam’ by allowing them to disclose their values through less extreme forms of hate speech expressed through moderate language.\textsuperscript{891} He cites the stirring up offences as an example. We have seen that these ban racist hate speech that is threatening, abusive or insulting and which is likely to or intended to stir up hatred. The law bends over backwards to allow citizens to disclose their racist values in moderate terms with even greater leeway granted to engage in hate speech towards religious groups.\textsuperscript{892}

Under the broader stirring up offences, a person could argue that Black people are underachieving in higher education because of their culture which promotes laziness, violence, and drug abuse. Similarly, Brown notes that under the narrower stirring up offences, one could claim, ‘I believe that Islam is a wicked faith and a danger to the finest institutions and traditions of this country’.\textsuperscript{893} They could go further and claim that Islam is a uniquely violent religion and that all of its followers should be viewed with

\begin{flushright}
\textsuperscript{889} ibid 164. \\
\textsuperscript{890} ibid 156. Baker, ‘Harm, Liberty, and Free Speech’ (n 36) 1019–1020; Waldron (n 1) 163; and R George Wright, ‘The Unnecessary Complexity of Free Speech Law and the Central Importance of Alternative Speech Channels’ (1989) 9 Pace Law Review 57. \\
\textsuperscript{891} Oster (n 24) 33. \\
\textsuperscript{892} The scope of the stirring up offences is set out in chapter 1 \\
\textsuperscript{893} Brown, \textit{Hate Speech Law} (n 55) 27.
\end{flushright}
suspicion. None of this would breach the law as we have seen that the narrower stirring up offences are limited to threatening material with an intent to stir up hatred. Post is right when he notes that 'It does not take clairvoyance to know that professional or even popular sociological articles analyzing Islamic terror cells will not be punished as hate speech. Nor will elite discussions of racial violence or the burdens of immigration'.\textsuperscript{894} Post presents this as a criticism. That hate speech laws are ineffective because they do not prevent the expression of all forms of bigotry. But this misses the point. The purpose of these laws is not to suppress all racist speech but rather to strike a balance between allowing all citizens—including racists—to disclose their values while protecting minorities from at least some extreme forms of hate speech.

Preventing the expression of all racist views will erode formal autonomy, while allowing a free for all for the expression of the most virulent racist views in public discourse will unduly harm minority groups. A balance must be struck between the two, and there is no avoiding this. The current line-drawing is necessary to create a 'safe-haven' allowing autonomous self-disclosure while, on the other hand, it protects minority groups from the worst forms of vile abuse and the harms which are more likely to be caused by overt hate speech, such as psychological harm as well as severe emotional distress and silencing.

b. Dworkin's and Weinstein's arguments from democracy: A citizen’s right to hateful political participation

There are various versions of the argument from democracy. Alexander Meiklejohn argued that a free speech principle could be drawn from our commitment to democracy. Under this theory, the voting public, as the 'ultimate source of political authority', must have access to all information and ideas to enable them to make decisions on public policy.\textsuperscript{895} I do not explore Meiklejohn’s account here. Many have pointed out that it is not a convincing account for a free speech principle because it is predicated on the idea of popular sovereignty, and this idea is incompatible with limiting the powers of that sovereign, including their powers to limit speech.\textsuperscript{896} A free

\textsuperscript{894} Hare (n 60) 135.
\textsuperscript{895} Meiklejohn (n 312) 255.
\textsuperscript{896} Barendt, Freedom of Speech (n 6) 18–19; Schauer, Free Speech (n 279) 41; Weinstein (n 871) 26.
speech principle must be able to justify protecting speech despite the majority wanting it suppressed, something that Meiklejohn’s account fails to do.

Instead, I focus on Dworkin’s more compelling anti-majoritarian argument from democracy, which has implications on the appropriate scope of hate speech laws. Like Baker, Dworkin’s account is underpinned by political legitimacy, which he suggests demands that the state treats each citizen with equal dignity by respecting them as a ‘free and equal member of the community’. 897 He argues that a fair democratic process requires, at a minimum, both a right to vote and a right to contribute to the formation of public opinion noting:

_a majority decision is not fair unless everyone has had a fair opportunity to express his or her attitudes or opinions or fears or tastes or presuppositions or prejudices or ideals, not just in the hope of influencing others (though that hope is crucially important), but also just to confirm his or her standing as a responsible agent in, rather than a passive victim of, collective action._ 898

On Dworkin’s view, the majority only has the moral right to pursue collectivist goals if they allow every citizen to try to dissuade them from doing so by dissenting, arguing and swaying them with their speech. 899 While Baker argues that hate speech bans imperil the legitimacy of the entire legal system, Dworkin’s claim is more specific. He argues that they imperil the political legitimacy of enforcing specific legislation against hate speakers. Weinstein, who develops a similar argument to Dworkin, explains this using the following analogy:

_if someone is barred from expressing a view on a proposed tax increase or whether the nation goes to war, or the country’s immigration policy, to that extent and with respect to that citizen, the government is no democracy but rather an illegitimate autocracy._ 890

---

898 ibid.
900 Weinstein (n 871) 29.
Dworkin and Weinstein would argue that these citizens have no moral duty to obey these laws and the state ought not to enforce them. Similarly, they argue that hate speech laws imperil the political legitimacy of anti-discrimination laws. Such laws are paramount as they prevent ‘specific and damaging consequences’ for minorities in ‘employment, education, housing’ etc.\(^{901}\) Racists vociferously oppose anti-discrimination laws. They would detest being forced to employ people from ‘lesser’ races that they deem to be incompetent. Islamophobes may not want to be forced to rent a house to Muslims because they believe them to be dangerous and would make the neighbourhoods unsafe. Homophobes may refuse to provide same-sex couples with services because they believe them to be degenerates and heretics. Dworkin and Weinstein argue that in order for society to have the necessary moral authority to enforce these anti-discrimination laws, they must not prevent these would-be dissenters from being able to express their opposition to them, to convince the majority that they are unnecessary and unjustified, however distasteful this form of opposition may be.\(^{902}\)

However, as with Baker’s argument for self-disclosure, the fact that Dworkin and Weinstein present a compelling argument for treating hate speech as high-value speech does not make the harms caused evaporate, and we must consider how it fits into a balancing approach. Firstly, political legitimacy is not an on/off valve but rather exists on a continuum. When we say that a law is illegitimate, what we mean is that a citizen has a right to disobey it, and the state ought not to punish them for doing so. In the words of Martin Luther King Jr:

"an individual who breaks a law that conscience tells him is unjust, and who willingly accepts the penalty of imprisonment in order to arouse the conscience of the community over its injustice, is in reality expressing the highest respect for law."\(^{903}\)

Martin Luther King Jr had no obligation to obey segregationist policies. Similarly, the inhabitants of Nazi Germany were also entitled to disobey laws that criminalised aiding Jews because these laws were unjust and, therefore, illegitimate. According to a

\(^{901}\) Weinstein (n 13) 529; Dworkin, ‘Foreword’ (n 897) 29.

\(^{902}\) Weinstein (n 13) 529; Dworkin, ‘Foreword’ (n 897) 28–29.

The racist’s conscience, laws preventing them from expressing their distaste for minorities are also unjust and illegitimate. They have no duty to obey them, and the state has no duty to enforce them against him. Further still, anti-discrimination and hate crime laws that protect minorities from assault and discrimination in areas such as employment and housing are equally unjust and illegitimate because the racist was not allowed to oppose them using his speech. By preventing him from arguing that Africans are of an inferior species of lower moral worth than himself—and are therefore undeserving of equal protection under the law—the state lost its right to enforce these laws against him. He can disobey them too. The police must stand down and let racists violate these laws. Hurling racial epithets, beating up minorities on the streets and refusing to hire them or let an apartment to them is as much an exercise of civil disobedience as Martin Luther King’s sit-in protest against segregationist laws.

I draw this vivid contrast only to point out that what is unjust and, therefore, illegitimate is not determined entirely by reference to the subjective conscience of the affected party. When we say that human beings are of equal moral worth and dignity, we mean that in civilised societies, it is no longer a matter of serious contestation that racial segregation is morally abhorrent and laws that seek to enforce it are profoundly unjust and illegitimate ab initio. We recognise that the laws Martin Luther King Jr violated were unquestionably more illegitimate than the ones Dworkin’s and Weinstein’s formulation suggests are delegitimised by hate speech laws.

Secondly, neither factor in that political legitimacy appears on both sides of the issue. We have seen how the failure to proscribe hate speech from individual hate speakers can harm vulnerable minority groups directly and indirectly by silencing them and preventing them from participating in democratic debate. As Gelber and Brown rightly point out, ‘the more that minority citizens are silenced or marginalized by hate speech…the less likely it is that they will identify with the state in the manner required by’ political legitimacy. As such, it is arguable that the failure to proscribe such speech also imperils the political legitimacy of enforcing laws against groups targeted by hate speech.

904 Waldron, The Harm in Hate Speech (n 14) 185.
Once it is conceded that legitimacy exists on a spectrum and that it appears on both sides of the hate speech debate, then Dworkin’s and Weinstein’s arguments can be profoundly useful in a balancing approach. As Waldron notes, it becomes a question of how great a deficit in legitimacy is created by a specific law:

*if the deficit is slight, then it may not generate a compelling case against hate speech laws when the stakes on the other side (the harms that such laws may avert) are very high.*  

906

It is also a question of the deficit in legitimacy created by the absence of such legislation. As with Baker’s argument from self-disclosure, an optimal balance between these conflicting values can be achieved by proscribing the most extreme forms of hate speech while allowing citizens to participate in democratic debate in more moderate terms. Bigots may be able to oppose the wearing of the burka in public by couching their bigotry as a concern for public safety, but they should not claim that women who wear face veils are filthy terrorists hiding bombs under their clothes.  

907 Such laws only create a slight deficit in the political legitimacy of enforcing antidiscrimination laws against hate speakers. At the same time, we have seen that overt hate speech is most likely to cause psychological harm, severe-emotional distress and silence minorities and exclude them from public discourse. By proscribing these forms, the state becomes more legitimate in the eyes of the protected minority groups.  

908

Under Baker’s as well as Dworkin’s and Weinstein’s accounts, hate speech is high-value speech that ought to benefit from strong protection and any restrictions placed on it—at least where lone hate speakers are concerned—should be limited to overt forms. The law must allow alternative modes of conveying the same message, such as through DNS, in order to maintain political legitimacy. Such accounts are difficult to reconcile with any broader restrictions on hate speech. The question, therefore, is whether the press can rely on these rationales.

906 Waldron, *The Harm in Hate Speech* (n 14) 188.  
907 Parekh, ‘Is There a Case for Banning Hate Speech?’ (n 569) 39.  
908 Brown, *Hate Speech Law* (n 55) 203.
c. The inapplicability of speaker-based accounts to the press

Baker’s argument from autonomy assumes that the hate speaker is disclosing their values when they engage in hate speech. Dworkin and Weinstein assume that hate speech is a citizen’s contribution to public debate. These assumptions do not apply in the case of the press, and these arguments are weaker or inapplicable to them. Rowbottom points out that the speech of mass media institutions does not manifest the values of a particular speaker but is instead the ‘product of interactions of various different components, such as multiple reporters, photographers, headline writers and editors’.909 This view was endorsed by Lord Leveson who noted that ‘press organisations are not human beings with a personal need to be able to self-express’.910

Some have criticised this view, arguing that whereas a press outlet may not be a human being, it is comprised of individuals who have speech rights.911 It is therefore argued that its output can be viewed as the autonomous expression of the values of the various human beings within the organisation and the contribution by these citizens to democratic debate. These arguments clearly have some applicability to citizen journalists and bloggers but are ill-suited to the mass media institutions such as the national or regional press that I am primarily concerned with. Whose values do the speech of these institutions embody? Is it the owners, editors, or journalists?

Journalists do not control what stories they pursue and write; they are subject to the direction of their editors on what they write and the stances they can take and operate under the terms of their contract of employment. As Rowbottom notes, ‘an analogy can be drawn with a teacher, who does not purport to exercise his or her own self-expression when speaking in the classroom. Instead, the teacher executes the task required by their employment contract.’912 Along the same lines, Oster argues that a journalist publishing content is not expressing their values but rather ‘acting in their professional capacity’.913 Further, even if the journalist’s views are aligned with the

---

911 Wragg, *A Free and Regulated Press* (n 18) 166.
913 Oster (n 24) 33.
editorial slant of the press outlet, this can be explained by a need to prevent cognitive dissonance.\textsuperscript{914}

Similarly, the editor’s relationship with a publication is also governed by a contract of employment, and owners who do not like an editorial line can direct it to be changed, or they can replace the editor. Then is it the owner’s free speech that is pertinent?\textsuperscript{915} The relationship between the media mogul Rupert Murdoch and his editors has often been described in these terms, with his papers said to reflect his worldview.\textsuperscript{916} A former editor of one of his papers, the \textit{Sunday Times}, wrote that ‘Rupert had an uncanny knack of being there even when he is not. When I did not hear from him...he was still uppermost in my mind’.\textsuperscript{917} However, other editors suggest that he is apolitical, and his stances were based both on furthering his business interests and publishing content that would sell.\textsuperscript{918} As an example, the Scottish edition of \textit{The Sun} newspaper often takes opposite stances on issues from the England and Wales edition. During the 2015 UK general election, they endorsed rival political parties.\textsuperscript{919} It is difficult to attribute the output of these contradictory views to the same owner, and it makes little sense to justify legal protection on the basis that a newspaper’s output reflects their values. Barendt is therefore correct when he notes that:

\begin{quote}
if we ask who can claim freedom of press within a newspaper, the usual answer is that ultimately the freedom must be ascribed to the press owner, for he is free to choose and fire an editor, who in his turn has authority to determine the contents of the paper. But that freedom can hardly be equated with freedom of expression; it is more akin to a property right or to commercial freedom.\textsuperscript{920}
\end{quote}

Baker excludes the institutional press from his self-disclosure argument on this basis.\textsuperscript{921} He argues that media institutions are corporations and corporate speech is not the ‘freely chosen expression of the speaker’.\textsuperscript{922} Rather, it is market-determined.

\textsuperscript{914} Baker, \textit{Human Liberty and Freedom of Speech} (n 237) 202.
\textsuperscript{915} Barendt, \textit{Freedom of Speech} (n 6) 441–442.
\textsuperscript{916} Wragg, \textit{A Free and Regulated Press} (n 18) 121.
\textsuperscript{917} Andrew Neil, \textit{Full Disclosure} (Pan 1997) 165.
\textsuperscript{918} Curran and Seaton (n 26) 204.
\textsuperscript{920} Barendt, ‘Statutory Underpinning: A Threat to Press Freedom?’ (n 167) 191.
\textsuperscript{921} Wragg, \textit{A Free and Regulated Press} (n 18) 69.
\textsuperscript{922} Baker, \textit{Human Liberty and Freedom of Speech} (n 237) 197.
In a free market economy, orientation toward anything other than profit makes a corporation non-competitive and will lead to its failure. The goal of profit maximisation is externally imposed by the market economy. As Baker notes, a corporate entity has no ‘general goal other than the instrumentally rational goal of increasing profits. Any other goal would detract from that rationality’.  

The press in the UK is, with a few exceptions, comprised of commercial institutions whose survival depends on selling papers, generating clicks and advertising revenue. This is not a novel observation, and critics of the institutional press have long claimed that market forces lead them to prioritise sensationalism, gossip, and violence over careful public interest journalism. Baker argues that content is ‘reduced to the lowest common denominator of the target market’. Similarly, Wragg claims that in the UK, the ‘race for market share’ determines ‘both content and editorial approach’. He notes that editors who testified before the Leveson inquiry identified their principal duty as selling papers and financial performance. Their speech is just as coerced as ordinary corporate speech, albeit with some qualitative differences. While the market dictates the message from both, a car dealer has a limited number of goods which he can advertise, whereas a press outlet’s ‘goods’ are its speech which allows it greater freedom of choice to present various ideas. Speech by a casino must exalt the virtues and benefits of gambling, and a coal company must promote the use and benefits of burning fossil fuels. They are forced to ‘lead listeners to one specific substantive conclusion’ whereas a newspaper can take either side on these issues leaving intact some individual judgement on content or viewpoint that is unavailable to other corporate speakers. Therefore although media products are also largely market determined, they are not just toasters with pictures.

Media law scholars have argued that the press also engages in hate speech to sell papers. Wragg points out that self-interest in generating profits often leads to

---

925 Baker, Human Liberty and Freedom of Speech (n 237) 226.
926 Wragg, A Free and Regulated Press (n 18) 69.
927 ibid 69–70.
928 Baker, Human Liberty and Freedom of Speech (n 237) 228.
discriminatory coverage against certain groups. Similarly, Rowbottom notes that speech that appeals to popular prejudices is ‘a type of hate speech that may still be acceptable to a large proportion of the population and such sensationalised stories can help to boost sales.’ Various other commentators have made similar observations. A former journalist from the Daily Star who testified before the Leveson inquiry stated that ‘editorial decisions are dictated more from the accounts and advertising departments than the newsroom floor’ and there was ‘a top-down pressure to unearth stories which fitted within a certain narrative (immigrants are taking over, Muslims are a threat to security).’

This form of ‘unethical but financially rewarding journalism’ is driven by a relentless pursuit of the ‘commercial agenda’. It has led to inflammatory fabricated stories about tax money being spent on ‘Muslim only toilets’ or Christian children being ‘forced into Muslim foster care’. Press hate speech then, is best understood as profit-motivated rather than value-disclosing speech.

As a rejoinder, Rowbottom argues that ‘if the true motives of speakers are examined to decide whether expressive acts are true statements of a person’s belief or values, then much speech is vulnerable to being viewed as “unfree”.’ He argues that speech is often as constrained and directed by social pressures such as a fear of offending friends, and there may not be a basis to distinguish such pressure from commercial ones. Further, he argues that commercial pressures do not apply to state-owned and charity-owned media, but these are still subject to a plethora of institutional constraints and are not merely vehicles for the heads of the public broadcaster or charity to self-express. Rowbottom is certainly correct here, but he understates the

---

930 Wragg, A Free and Regulated Press (n 18) 70, 176.
931 Hare (n 60) 627.
932 Various commentators have noted that the press write sensationalised stories about ethnic minorities to boost sales Hare, Extreme Speech and Democracy (n 53) 628; Memorandum submitted by the PCC to the Joint Committee on Human Rights, The Treatment of Asylum Seekers, Vol. 2, HL Paper 81-II, HC 60-II, Tenth Report of Session 2006–2007, 445–6.
933 Wragg, A Free and Regulated Press (n 18) 176.
934 ibid.
935 ibid.
936 Rowbottom, Media Law (n 42) 10.
937 ibid.
938 ibid 9.
commercial pressures that even public broadcasters are subject to, as John Parkinson points out:

the media in modern democracies are embedded in a political economy in which information and cultural production are more or less marketized: even those media which are largely state supported still face some competitive pressures to attract those audiences which in turn attract revenues.\(^{939}\)

He uses the example of the BBC, whose funding historically has been tied to its ability to attract audiences.\(^{940}\) Rowbottom also arguably misreads Baker. The latter is not suggesting a deep inquiry into whether people really believe what they say before protecting their speech. His concept of autonomy is entirely compatible with a speaker being motivated by other forms of self-interest, noting that:

a person’s autonomy might reasonably be conceived as her capacity to pursue successfully the life she endorses—self authored at least in the sense that, no matter how her image of a meaningful life originates [emphasis added], she now can endorse that life for reasons that she accepts.\(^{941}\)

Rather, Baker argues that a profit motivation which is externally imposed by a market economy is coercive. In the previous chapter, I argued that social pressure can be powerful, but even this leaves a person’s formal autonomy intact. The speaker is still free to decide whether to offend their friends, whereas a corporation’s speech must succumb to a profit motive to survive. This differentiates commercial constraints from a speaker’s other self-interests. It is for these reasons that the overwhelming position endorsed in the SRT literature that speaker-based accounts for free speech are inapplicable to the institutional press is justified.\(^{942}\)

Press hate speech is therefore unsupported by Baker’s argument from autonomy. Equally, the other persuasive account explored above (a citizen’s right to political


\(^{940}\) ibid.


\(^{942}\) Scholars who express this view include Barendt, *Freedom of Speech* (n 6) 417; Oster (n 24) 141; O’Neill (n 267); Rowbottom, *Media Law* (n 42) 7–8; Damian Tambini, ‘A Theory of Media Freedom’ (2021) 13 Journal of Media Law 135, 149; Fenwick and Phillipson (n 20) 21–23.
participation) is inapplicable as a press outlet is an institution, not a voting citizen and does not possess political rights. Regulating DNS in the press will not erode political legitimacy as the individual journalists are free to take part in political debate in their personal capacity where they can exercise the same rights to engage in hate speech as the rest of the public without their speech benefitting from such institutional amplification. In short, press hate speech is not high-value speech under either of these arguments. Granting them less leeway to engage in hate speech, i.e. by proscribing DNS, is entirely compatible with both accounts.

4. A Non-instrumentalist audience-based account: Scanlon’s argument from audience autonomy

Even if the press cannot rely on speaker-based non-instrumentalist arguments to publish hate speech, they may be able to rely on free speech arguments that focus on the value that speech provides to its audience. These accounts are either instrumental or non-instrumental. I will examine each and argue that unlike the two speaker-based accounts explored above, they do not present a compelling case for treating hate speech as high-value speech under a free speech principle. I will explore three accounts, starting with Scanlon’s argument from audience autonomy.

Scanlon appears to have abandoned his position in later work. However, it has been favourably cited by Schauer, Barendt and Dworkin. More importantly, Wragg has relied on a similar account to argue that content restrictions on the press violate the audience’s autonomy. This argument and its implications on press hate speech merit exploration. Like Baker, Dworkin and Weinstein, Scanlon argues that restrictions on free speech imperil political legitimacy. Unlike them, he focuses on the right to receive information rather than the right to impart it. He argues that a legitimate state must allow individuals to have sovereignty over their minds. States must not restrict the information they can receive as this takes away their autonomy to determine the

944 Schauer, Free Speech (n 279) 70.
947 Wragg, A Free and Regulated Press (n 140) ch 5.
948 Scanlon (n 290) 215.
merits of the ideas and decide for themselves what actions to take. Any such restrictions are illegitimate because they fail to treat citizens as ‘equal autonomous and rational agents’. This position broadly mirrors the argument developed by Mill in the second chapter of *On Liberty*, and Scanlon refers to his theory as a ‘Millian principle’. It rests on liberal notions of individual agency by assigning the responsibility for coming to believe and acting on ideas on the individual (who holds and acts on those ideas), even if that leads to bad outcomes, as the only alternative is to surrender ‘independent judgement to the state’. Scanlon has associated his theory with John Rawls’s hypothetical social contract arguing that those in the Original Position would not grant the authority to the state to ‘decide on matters of moral, religious or philosophical doctrine (or of scientific truth)’. Wragg opposes content-based press regulation on this basis. He argues that the press is merely delivering what people want, and the market share of the tabloids reflects the reader’s personal desire for ‘celebrity gossip or reactionary diatribes’. Consequently, even if such speech does not reflect the speaker’s autonomy, it reflects the audience’s autonomy, which would be violated when they are denied the opportunity to consume content that they are clearly interested in. Even assuming that this is a persuasive account of a free speech principle, there are reasons to be doubtful as to whether the content that sells the most does indeed reflect audience preferences.

Firstly, we could easily flip the script here and argue that it is the publication of such content by the press which creates these preferences in the first place. Recall how the media is a principal source of the original encoding of stereotypes in our memories and increases their availability through priming. It could be argued that the press is selling the prejudice that they fomented themselves, and even the consumption of

949 Schauer, *Free Speech* (n 279) 69.
950 Scanlon (n 290) 215.
951 ibid 213.
952 Barendt, *Freedom of Speech* (n 6) 17.
954 Schauer, *Free Speech* (n 279) 70; Brison (n 346) 328.
956 See discussions in chapters 1 and 5.

177
such content does not reflect the autonomous choices of the audience. It is on this basis that Rowbottom argues that imposing restrictions on the press can be justified because they do not ‘merely represent the preferences of the audience but shapes its views’.  

Secondly, Wragg assumes here that the reason why such content is published by various outlets reflects audience preferences rather than the outlets maximising efficiency. The greatest profits are to be found in publishing content that costs the least to produce, and celebrity gossip, as well as sensationalised stories about ethnic minorities and asylum seekers, are far less costly in terms of production and legal costs than investigative journalism as marginalised groups are often not in a financial position to pursue claims against the media. The latter may be far more valuable to and preferred by the audience, but the market disincentivises its production. This leads Fiss to argue that by maximising profits and reducing costs, the press offers the public considerably less than they want.

More importantly, there are good reasons why Scanlon abandoned this account as it is not a convincing rationale for a free speech principle. Firstly, as many have pointed out, restrictions on speech are perfectly consistent with audience autonomy if the audience itself wants the state to restrict access to certain information and ideas. Greenawalt notes that rational, autonomous people in the Original Position may actually want certain speech restricted because it is ‘too dangerous to be tolerated’. They may rather ‘foreclose some inputs’ in order to prevent irrational people from receiving and acting on them. Secondly, Barendt argues even assuming that people are rational, there are good reasons to doubt whether some forms of speech, such as hate speech, ‘will be considered rationally by its audience’. Rowbottom notes that media speech influences through more indirect effects, such as framing rather than

---

957 Hare (n 60) 628.
960 Rowbottom, Media Law (n 42) 17.
961 Barendt, Freedom of Speech (n 6) 17.
962 Greenawalt (n 287) 115.
963 Barendt, Freedom of Speech (n 6) 33.
rational persuasion. In earlier chapters, I have set out at great length how Barendt and Rowbottom’s views here are correct. They are supported by empirical evidence from cognitive, social psychology, behavioural economics and media effects which suggests that media stereotyping changes attitudes towards minorities in subtle, indirect, and non-rational ways. I have also argued that in certain circumstances, the media has the necessary epistemic authority which provides strong reasons to update one’s beliefs. It is then arguable that restrictions on such speech enhance rather than violate audience autonomy. Finally, it is also plausible that those in a hypothetical Original Position would allow restrictions on hate speech because they would not want to take the risk that they would be members of a minority group and be victims of it. Consequently, this account does not support treating press hate speech as high-value speech and does not pose a challenge to regulating DNS.

5. An Instrumentalist audience-based account: The argument from self-fulfilment

I will now briefly consider one instrumentalist audience-based argument for a free speech principle appearing most prominently in the work of Martin Redish, this being the argument from self-fulfilment. Under this theory, speech is seen as a core part of the development of personality and intellectual growth. It is argued that restricting what we are allowed to say, write, read or hear ‘inhibits our personality and its growth’. Mill is another who puts much stock in notions of mental self-development and self-reflection, arguing that ‘the human mind’ is the ‘source of everything respectable about a man, either as an intellectual or a moral being’. It is argued that by communicating, reading and hearing from others, we become ‘more reflective and mature individuals’. This argument then supports press speech claims in so far as

964 Rowbottom, Media Law (n 42) 17.
965 Barendt, Freedom of Speech (n 6) 17; Brison (n 346) 328.
966 Brison (n 346) 329.
967 Redish refers to it as the argument from Self-realization but the more common usage is self-fulfilment and I use that term here see Barendt, Freedom of Speech (n 6) 13; Schauer, Free Speech (n 279); Martin H Redish, ‘Value of Free Speech’ (1982) 130 University of Pennsylvania Law Review 591.
968 Barendt, Freedom of Speech (n 6) 13.
969 Mill (n 287) 34–35.
970 Barendt, Freedom of Speech (n 6) 13.
it suggests that reading and hearing different information and ideas allows the 'development of one’s human faculties'.

As a general point, instrumentalist justifications collapse if the claimed benefit does not arise from the speech. While it is certainly plausible that some people will achieve personal happiness and growth from expressing their opinions and hearing the views of others, many may not. Some may, as Schauer puts it, ‘prefer the security or intellectual anaesthesia that accompanies rigid controls on expression.’ Similarly, a racist may draw immense pleasure from reading an article in a national daily containing DNS about groups, but the unrestricted circulation of such content may even hinder the happiness and personal growth of the vast majority of non-racists. The instrumental value in receiving these ideas is therefore doubtful, while on the other hand, we have evidence that such speech will inhibit the self-fulfilment of the groups it targets. I have set out the deleterious effects DNS have on a target group’s self-esteem and the esteem in which others hold them and that this limits their ability to lead fulfilling lives. These are all arguably more damaging to a person’s self-fulfilment than any personal growth attained from the circulation of such speech. This argument, therefore, supports, rather than limits, the broadening of hate speech restrictions.

However, the weakest aspect of this account, as many have pointed out, is that it does not explain why speech is uniquely important for self-fulfilment. It does not make clear why intellectual development is more important than other human needs such as food and personal security, shelter, as well as physical and emotional well-being so as to justify special protection. Robert Bork puts it best when he notes that an individual may achieve personal growth and happiness from ‘trading on the stock market, following his profession as a river-port pilot, working as a barmaid, engaging in sexual activities’ and the self-fulfilment rationale fails to distinguish speech from ‘all other human activity’. It is more akin to Baker’s notion of substantive autonomy which is enhanced by the possession of material resources. The enhancement of

971 Redish (n 967) 627.
972 Brison (n 346) 319–320.
973 Schauer, Free Speech (n 279) 49.
974 Barendt, Freedom of Speech (n 6) 13; Schauer, Free Speech (n 279) 56.
975 Barendt, Freedom of Speech (n 6) 13; Schauer, Free Speech (n 279) 56.
every person’s substantive autonomy ought to be a primary goal pursued by a state, but it does not operate as a normative constraint on the exercise of power. A claim to a free speech right that is of equal footing to a non-exhaustive list of wants and desires is not a justification for a free speech principle. Further, it is arguable that this account is better suited to justify protecting these other non-speech activities because they do not create the same negative externalities caused by hate speech. Consequently, while this account certainly applies to press hate speech, it is not a persuasive argument for a free speech principle.

6. Conclusion

This chapter explores whether imposing broader regulation on press hate speech that goes beyond those imposed on the public is compatible with SRT. I have answered this question in the affirmative by demonstrating that the most persuasive rationales for treating hate speech as high-value speech are two non-instrumentalist speaker-based accounts offered by Baker, Dworkin and Weinstein, neither of which apply to the press.

Both accounts are grounded on political legitimacy. In Baker, we find a persuasive rationale explaining why it demands respect for formal autonomy, and this requires leeway to be granted to hate speakers to disclose their hateful beliefs to others. In Dworkin and Weinstein, we find a compelling argument as to why political legitimacy requires that hate speakers be allowed to participate in political debate and voice their opposition to the equal status of minority groups in society. Both require treating hate speech as high-value speech, and in a balancing approach, this means only regulating overt hate speech and granting speakers the leeway to disclose their values and participate in political debate through more moderate language. However, these arguments do not apply to the press. Their hate speech is profit-motivated, not value disclosing speech and is either low or no-value speech under a free speech principle. Broader restrictions of their hate speech (including proscribing DNS) are entirely compatible with a free speech principle.

977 Schauer, Free Speech (n 279) 56.
This only answers half of the SRT challenge to such regulation; in line with the second condition (see section 2 above), it must also be demonstrated that restricting DNS does not hinder the press’s ability to perform their democratic functions. I will consider this in the next chapter.
Chapter 8: Social Responsibility Theories and Press Freedom

1. Introduction

The previous chapter demonstrated that broader restrictions on press hate speech are compatible with a free speech principle. I will now consider whether such restrictions are compatible with a principle of press freedom, understood here as encompassing the various democratic functions assigned to the press in SRT.

These functions vary in breadth and scope based on the model of democracy, but three core functions are identified in SRT literature, and I consider each below. Section 2 briefly recaps what is covered and protected by a principle of press freedom and how it ought to be balanced against competing interests. Section 3 explores the watchdog role, setting out the limited circumstances under which it requires the media to publish DNS. Section 4 considers different democratic models and how they conceptualise the media’s role in imparting information and ideas and considers whether and to what extent they include a right to impart hateful information and ideas. Section 5 engages in the same analysis for the platforming function.

2. A principle of press freedom: coverage and protection

Recall as with a free speech principle, certain activities in newsgathering and publishing are not covered by a principle of press freedom. For instance, it was noted that activities such as phone hacking or a media outlet renovating its offices are not covered by this principle (no-value activities). Restrictions on no-value activities through, say, enforcing building regulations do not implicate the principle. There is no need for balancing, and the state has wide discretion to restrict them to prevent harm.

Low-value activities are those which are covered by the principle but lie at its margin. I have used celebrity gossip as an example. Low-value activities benefit from minimal protection and are accorded less weight when balanced against the need to prevent the various harms caused and constituted by press hate speech. Optimal balancing in

---

978 Oster refers to this as a Media Freedom principle, a term which he borrows from Schauer’s idea of free speech principle explored in chapter 2 Oster (n 24) 2.
such cases requires prioritising the principle of harm prevention by regulating the low-value activities (publication of DNS) causing these harms.

On the other hand, activities which are at the core of the principle—those which fulfil certain democratic functions assigned to the press—are covered and benefit from the greatest protection. These would be high-value activities that ought to benefit from special immunity from regulation as well as the special privileges necessary for the press to fulfil those functions, e.g. protection of sources, protection from surveillance, access to information etc.\textsuperscript{979} High-value activities can only be subject to narrow restrictions such as those found in the stirring up offences. It will therefore be difficult to reconcile restrictions on DNS if press hate speech is a high-value activity.

I will argue that in most cases, hate speech is either low or no-value for purposes of a principle of press freedom as publishing such content is rarely required for the press to fulfil its various democratic functions. If this is correct, regulating DNS is compatible with press freedom. I will also argue that such regulation can cater for the limited instances where press hate speech is a high-value activity by exempting the press from liability in such cases.

3. The watchdog role and hate speech

A core democratic function assigned to the press is that of a public watchdog.\textsuperscript{980} They are the ‘eyes and ears of the general public’ investigating and reporting abuses of power.\textsuperscript{981} This role envisions the press as a ‘fourth estate’ and sees them as—alongside the judiciary and parliament—a vital part of the system of checks and balances.\textsuperscript{982} Rowbottom notes that ‘the defining image of the modern watchdog function is reflected in the Watergate scandal, in which the journalist is presented as a crusader uncovering the abuse of high-level power’.\textsuperscript{983} The watchdog function is also preventative, as the knowledge that one’s wrongdoing may be exposed by the press

\textsuperscript{979} For an overview of different special privileges see ibid 85–101.
\textsuperscript{980} Blasi (n 235) 542; Rowbottom, \textit{Media Law} (n 42) 22.
\textsuperscript{981} Barendt, \textit{Freedom of Speech} (n 6) 418.
\textsuperscript{982} Baker, \textit{Human Liberty and Freedom of Speech} (n 237) 230.; It is emphasised as the basis for press freedom in the UK by the Third Royal Commission on the Press see Barendt, \textit{Freedom of Speech} (n 6) 420.
\textsuperscript{983} Rowbottom, \textit{Media Law} (n 42) 20.
discourages misconduct in public office.964 This function is not limited to public offices as the press equally disincentivises and exposes corporate corruption.985 The watchdog role is recognised and emphasised by the UK courts and the ECtHR in several cases.986 However, the courts adopt a broader conception of this function to include a duty to impart any information in the public interest. I will consider this as a separate function below. Here I focus on the narrower conception which sees it as encompassing any activities which involve investigating and reporting corruption, incompetence, or abuse of power. There are two aspects of the watchdog role that inform both the scope and structure of any regulation of DNS.

In terms of scope, this function is at the core of press freedom and ought to benefit from the strongest protection.987 This means that it is difficult to justify the regulation of DNS when the press is performing it. This is the position taken by the ECtHR which subjects restrictions that impinge on the press’s ability to perform this function to heightened scrutiny, with states having only a limited margin of appreciation.988 However, this is largely unproblematic because, in most cases, exposing government wrongdoing or holding power to account does not require the press to engage in hate speech. There are certainly limited circumstances where this function requires the press to report on matters which may inadverently promote certain DNS against groups, in which case such activities are covered by the principle and benefit from strong protection. An example of this is the Rotherham Child Sexual exploitation scandal, where it was alleged that the police failed to seriously investigate and prosecute child sex abuse in part because the perpetrators were mostly men from an ethnic minority background, and the police feared being perceived as racist and inflaming inter-communal tensions.989 Undoubtedly, such reporting may lead to irrational hatred against people from a similar ethnic background as the perpetrators,

985 Baker, Media, Markets, and Democracy (n 958) 133.
986 Reynolds v Times Newspapers Limited and Others [2001] 2 AC 127; Sunday Times v UK (No.2) [1979] Application number 13166/87; Axel Springer AG v Germany (No. 1) [2012] Application number 39954/08, para 79; Times Newspapers Ltd v United Kingdom (Nos. 1 and 2) [2009] Applications nos. 3002/03 and 23676/03, para 40.
987 Barendt, Freedom of Speech (n 6) 417–418; Oster (n 24) 1, 40.
but restricting it would be unjustified as it would unduly limit and erode the press’s ability to discharge its watchdog role. Rowbottom notes that regulation in such cases ‘should be framed in a way that preserves the scope for the media to investigate the government and other powerful institutions.’ 990 Similarly, Oster points out that in the balancing exercise between press freedom and restrictions on hate speech, relevant factors weighing in favour of press freedom are whether the publication is accurate and in the public interest. 991 This suggests that any regulation of DNS in the press ought to exempt them from liability where they publish accurate statements of fact that expose wrongdoing, abuse of office or incompetence.

The requirement of accuracy matters because fabricated stories that promote hatred do not further the watchdog role. For example, in 2017, The Times published a story headlined ‘Christian child forced into Muslim foster care’. 992 It suggested, in provocative terms, that a ‘white Christian child’ was taken from their family by police and placed into emergency foster care with a devout Muslim carer, one of whom wore a face veil. The carers allegedly did not speak English and forced the girl to remove the crucifix she was wearing, stopped her from eating bacon and encouraged her to learn Arabic. 993 The problem here was that this story was riddled with inaccuracies. The supposedly ‘white Christian’ child was of Muslim heritage and was taken from her mother by the police because of the mother’s alleged alcoholism and drug usage. 994 The foster family spoke English, none of them wore a veil, and a court-appointed guardian found that the child was ‘settled and well cared for’ in the home. 995 In light of the facts, it is hard to see how this was even a matter of public concern, let alone an example of the press acting as a watchdog. At the same time, this reporting perpetuated DNS about Muslims, unjustifiably portraying the carers as holding extremist and intolerant religious views. 996 It unsurprisingly led to uproar, calls for a

990 Hare (n 60) 610.
991 Oster (n 24) 234.
992 Tower Hamlets Borough Council v The Times [2018] IPSO 20480-17
995 Grierson (n 993); Cathcart and French (n 994) 29–30.
996 Cathcart and French (n 994) 30.
public inquiry and was utilised by far-right groups to incite hatred against Muslims. This demonstrates the importance of making any watchdog exemption to DNS regulation conditional on accuracy.

However, the watchdog function also informs us about the structure of press regulation. We have seen that Baker argues that it prevents any regulation of the press by the legislative branch as this allows the potential abuser to set the rules for those exposing abuses. Similarly, Oster argues that the legislative branch should not determine the scope of press freedom because of the media’s role in scrutinising their behaviour and that the ‘independence of the media from state institutions is thus a requirement of democracy itself’. This limitation supposedly imposed by the watchdog role is a powerful rhetorical tool that has been relied on by the press to successfully campaign against any form of statutory regulation. In his evidence before the Leveson Inquiry, the editor of the Daily Mail claimed that any involvement by parliament in press regulation was the ‘thin edge of the wedge’ that would lead to statutory control. The Prime Minister expressed similar sentiments arguing that any form of statutory regulation would ‘cross the Rubicon’, meaning a ‘dreadful step which no liberal politician should ever contemplate’.

If the view that the watchdog function requires immunity from any form of statutory regulation is justified, it would be impossible to square DNS regulation with press freedom. However, there are several reasons why this view is mistaken. Firstly, this view of press freedom is dangerous as such immunity can be used by owners to serve their economic interests and to run roughshod over people’s rights without any accountability. Leveson rightly noted that ‘in a modern democracy that abides by the rule of law press freedom can never mean a press which sits outside, above and beyond, or in disregard of, the law’. Therefore, he rejected this view, and the

997 Wragg, A Free and Regulated Press (n 18) 178.
998 Oster (n 24) 234.
999 Baker, Human Liberty and Freedom of Speech (n 237) 231.
1000 Oster (n 24) 85.
1001 Wragg, A Free and Regulated Press (n 18) 264.
1002 HC Deb, vol 554, col 449 (29 November 2012).
1005 Leveson (n 2) 65.
primary and alternative recommendations by the inquiry envisioned some form of statutory underpinning to press regulation.¹⁰⁰⁶

Secondly, states such as Germany and Finland have enacted press laws containing accuracy requirements and rights of reply without leading to censorship or state control.¹⁰⁰⁷ Barendt points out that arguments about statutory restrictions being incompatible with press freedom ‘would be considered risible in Germany’ and ‘there seems to be no good reason why they should be treated more seriously in the United Kingdom’.¹⁰⁰⁸ I would add that the Danish Press Council operates a system of co-regulation established by statute, membership is mandatory, and it regulates both broadcasting and print.¹⁰⁰⁹ These are not authoritarian states. They all consistently rank higher than the UK on the Reporters Sans Frontieres Press Freedom Index. This index measures the degree of press freedom in different countries based on criteria such as independence from political actors, censorship, and access to information, all of which are central to the watchdog function.¹⁰¹⁰ While it is not a gold standard, it at least suggests that we ought to treat claims that deem any regulation of the press as an erosion of press freedom with a degree of scepticism.

Thirdly, this claim conflates activities that are covered by the principle and those which are not. The press is already subject to numerous restrictions on their activities. These include statutory and common law restrictions on contempt of court, as well as automatic and discretionary reporting restrictions that limit what can be published about criminal and civil proceedings.¹⁰¹¹ Most do not impact press freedom. The press is not acting as a watchdog when it discloses the identity of victims of sexual assault. We recognise that such activities are uncovered by press freedom and ought to be curtailed because they unjustifiably infringe on another’s rights.¹⁰¹² It would also be a perversion of the concept if press freedom was used to justify immunity from competition law.¹⁰¹³ I have argued that any regulations impinging on press activities

¹⁰⁰⁶ ibid 1782.
¹⁰⁰⁷ Barendt, ‘Statutory Underpinning: A Threat to Press Freedom?’ (n 167) 196; Leveson (n 2) 1721.
¹⁰⁰⁹ Leveson (n 2) 1718–1719.
¹⁰¹¹ Leveson (n 2) 66; Barendt, ‘Statutory Underpinning: A Threat to Press Freedom?’ (n 167) 195.
¹⁰¹² Oster (n 24) 37.
¹⁰¹³ Barendt, Freedom of Speech (n 6) 430.
that are uncovered by the principle are merely restrictions on commercial freedom and ought to be left to the discretion of lawmakers to determine.

However, what the watchdog function does require is that any restriction should be structured in a way that insulates the press from direct state interference, as this would compromise their ability to investigate abuses of power. I relegate the discussion of what is the most appropriate form of regulation to the next chapter. Here I only point out that any regulatory system must be independent both from the industry and from the government.\textsuperscript{1014} Regulation that allows for direct control of the press is \textit{prima facie} incompatible with the watchdog function. Barendt rightly points out that statutory regulation which creates a regulatory body packed with government appointees would be illegitimate, as would a return to a 17th-century licensing system.\textsuperscript{1015}

Oster also argues that criminal sanctions on activities covered by a principle of press freedom are ‘fundamentally contrary’ to and ‘destroy the essence of the right’.\textsuperscript{1016} I agree. Journalists must observe and obey the criminal law applicable to the rest of the public—including hate speech laws—and press freedom does not exempt them from this.\textsuperscript{1017} However, a press-specific criminal restriction that is broader than that which applies to the public would indeed be the thin end of the wedge, as it creates an avenue for direct interference with and intimidation of disfavoured publications by the government. Much has been written about the chilling effect on investigative journalism caused by criminal prohibitions targeted at—or with special application to—the press contained in official secrets and anti-terrorism legislation, and this should not be replicated here.\textsuperscript{1018}

Even if this were not the case, criminal law is a blunt instrument.\textsuperscript{1019} Recall that the content of stereotypes is highly group-specific. Antisemitic DNS will be different from Islamophobic ones. We have also seen that DNS against the same group manifests in subtly different ways and utilises different tropes depending on the society. These

\textsuperscript{1014} Oster (n 24) 38.
\textsuperscript{1015} Barendt, ‘Statutory Underpinning: A Threat to Press Freedom?’ (n 167) 197.
\textsuperscript{1016} Oster (n 24) 128, 138–139.
\textsuperscript{1017} Stoll v Switzerland [2007] (Application number 69698/01), para 102.
\textsuperscript{1019} Hare (n 60) 618.
tropes are memetic in that they evolve, change, and spread within a culture. What is required therefore is regulation that recognises these distinctions between groups and which is adaptable and can keep up with these cultural shifts. A non-contextualised and broadly formulated criminal provision is both ill-suited to this task and likely incompatible with the watchdog function.

4. Imparting hateful information and ideas

It is argued that the media has a unique role in imparting information and ideas to the public due to their unique capacity to ‘collect, filter, verify, package and disseminate information.’ While most normative accounts of democracy assign this role, some are more demanding than others. Here I look at some common accounts in the SRT literature centred around popular participation in democratic debate, which, as noted earlier, is viewed as one of democracy’s legitimating conditions. All have different visions of what popular participation entails and thus, what role the media ought to play in facilitating it. This is not to suggest that democracies, as they exist in practice, neatly map into any of these conceptions, and states may apply different principles to different media. However, I will argue that broadly, each of these models is compatible with some restrictions on DNS.

a. Republican and deliberative democratic models

The first model is what Baker refers to as the ‘Republican Model’, which sees democracy as a collective effort by the citizenry to discuss, formulate, commit to, and pursue common ends. These republican values of civic virtue, dialogue and consensus building also inform theories of deliberative democracy whose prominent exponents include Jürgen Habermas and, more recently, Robert Post. Deliberative

---

1021 Rowbottom (n 155) 18.
1022 Clifford G Christians and others, *Normative Theories of the Media: Journalism in Democratic Societies* (University of Illinois Press 2010) 95; Hare (n 60) 617–618.
democrats do not believe that public deliberation is only about pursuing common ends, but rather it also enables ‘ethical self-understanding and self-constitution’.\textsuperscript{1025} To deliberative democrats, democracy is legitimated by citizens accessing and communicating in the public sphere.\textsuperscript{1026} Oster, who develops a Habermasian account of press freedom, asserts that the media is the ‘institutional core of the public sphere’ whose duty is to ‘communicate information and ideas about political, cultural and other public issues’.\textsuperscript{1027} Thus, the democratic role of the media is both to provide the public with the information necessary to engage in rational deliberation and a platform to exchange ideas, enabling them to form a consensus on various issues. I will consider platforming as a distinct role below. Here, I focus only on imparting information and ideas. Some suggest that this entails impartial coverage of political affairs, akin to the public service model applied in broadcasting,\textsuperscript{1028} but deliberative democrats are primarily concerned with exposing citizens to the range of views held within a society, and this does not require giving equal weight to every view.\textsuperscript{1029} The question is whether this role entails granting leeway to the media to disseminate \textit{their own} racist views. There are two reasons to believe this is not the case.

Firstly, this model requires the media to be factually accurate for citizens to effectively participate in democratic debate and arrive at the truth. Oster argues that ‘ideal public discourse presupposes the best available information and arguments’ and to facilitate this discourse, the media ought to—and can be legally mandated to—provide citizens with factually accurate information.\textsuperscript{1030} This is a common position in SRT literature. Fenwick and Phillipson note that allowing the press to publish false information would mean citizens are participating in democracy from a ‘flawed perspective’, and the truth is unlikely to ever emerge in open debate if they do not have access to trusted sources of information.\textsuperscript{1031} This view is compatible with restrictions on DNS that are presented as fact. Recall that DNS are always false in so far as they treat people not as

\begin{thebibliography}{999}
\bibitem{1025} Baker, \textit{Media, Markets, and Democracy} (n 958) 147.
\bibitem{1026} The public sphere is a sociological term originating from Habermas where citizens’ gather to form public opinion Habermas, \textit{The Structural Transformation of the Public Sphere} (n 1024) ch 2; Post, ‘Participatory Democracy and Free Speech’ (n 1024) 482.
\bibitem{1027} Oster (n 24) 30.
\bibitem{1028} Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 614–615.
\bibitem{1029} Ibid 615; Baker, \textit{Media, Markets, and Democracy} (n 958) 148–149.
\bibitem{1030} Oster (n 24) 44–45.
\bibitem{1031} Fenwick and Phillipson (n 20) 31.
\end{thebibliography}
individuals but as a ‘uniform specimen’ of a racial or religious group, and they are therefore always inaccurate.\textsuperscript{1032}

This is a distinctly anti-Millian approach to truth-seeking. Mill believed that truth is best discovered through an open exchange of ideas. He argued that the expression of ideas we currently understand as being false ought not to be barred as they may be true. To prevent their expression is to assume that we are infallible.\textsuperscript{1033} Even ideas that we currently understand to be false may contain a grain of truth. Lastly, he argues that even if the ideas are completely false, they still serve a useful purpose as they challenge us, allowing us to hone the true ideas and prevent them from becoming a ‘dead dogma’.\textsuperscript{1034} A Millian approach to truth-seeking would therefore be incompatible with restrictions on DNS. To Mill, the false ideas that certain groups are dangerous or unintelligent ought to be defeated not by restrictions but by an open exchange in the marketplace of ideas.

My view is that the first approach is preferable to Mill’s as his argument rests on shaky assumptions.\textsuperscript{1035} The first assumption is that truth is best discovered through an open exchange of ideas. If objective truth exists,\textsuperscript{1036} it is not clear why entertaining patently false ideas enables us to get closer to it as ‘a needle is harder to find in a haystack than in two pieces of hay’.\textsuperscript{1037} As O’Neill argues, the discovery of truth requires a commitment to the norms of truth-seeking such as accuracy, ‘norms for using evidence and other everyday epistemic norms’.\textsuperscript{1038} She observes that in truth-seeking institutions such as universities, speech is highly regulated.\textsuperscript{1039}

The second assumption is that humans are rational beings who can parse out fact from fiction. Therefore, false information spread by the media will be detected and

\textsuperscript{1032} Parekh, ‘Is There a Case for Banning Hate Speech?’ (n 569) 44.
\textsuperscript{1033} Christopher T Wonnell, 'Truth and the Marketplace of Ideas' 19 U.C. Davis Law Review 669, 671.
\textsuperscript{1034} Mill (n 287) 58.
\textsuperscript{1035} Ba\textsuperscript{ker}, Human Liberty and Freedom of Speech (n 237) 12. They are many other criticisms of Mill’s argument see Robert C Post, Democracy, Expertise, and Academic Freedom: A First Amendment Jurisprudence for the Modern State (Yale University Press 2012) 4.
\textsuperscript{1038} M Madianou and A Pinchevski, Ethics of Media (N Couldry ed, Palgrave Macmillan UK 2013) 24.
\textsuperscript{1039} ibid 24.
rejected by the public. In the context of hate speech, this would mean that rational beings would reject the irrational and false stereotypes that the press publishes about various groups. However, we have seen that the social sciences suggest otherwise, with our desire for truth tempered by the need to conserve cognitive resources, bounded rationality, cognitive dissonance, and motivated reasoning. Parekh also argues that the marketplace of ideas ‘operates against the background of prevailing prejudices. When racist, anti-Semitic, and xenophobic beliefs are an integral part of a society’s culture, they appear self-evident, common-sense, and obvious, and therefore enjoy a built-in advantage over their opposites’.1040 Ideas that certain groups are inferior to others have been actively traded and have survived in the marketplace for generations despite counter-speech.1041 Many have therefore concluded that Mill’s reliance on ‘the rationality of the human mind’ has not been justified.1042

Moreover, the marketplace of ideas has no obvious corrective measure when irrationality prevails, and people choose the ‘bad ideas’ of hate speech after listening to all the propositions.1043 Oster is correct when he notes that if truth-seeking is the goal, press freedom does not require providing the media with the room to perpetuate the false ideas of holocaust denial, and I would add DNS to this list.1044 Such accuracy requirements would certainly address and cover some forms of press hate speech. For example, the accuracy clause in the Editors Code’ has been used to successfully obtain corrections on stories that perpetuate stereotypes about groups in some cases.1045

However, accuracy requirements have limited utility for two reasons. Firstly, it is unrealistic to impose stringent accuracy requirements on the press. In elaborate disciplines such as academic research, truth claims are subject to verification through processes such as peer review that allow them to be rigorously tested before publication. News, on the other hand, is a perishable commodity and has quick turnaround times. Moreover, journalists, unlike academics, often lack the expertise in

---

1041 Wonnell (n 1033) 671.
1042 Ingber (n 1036) 7.
1044 Oster (n 24) 228–229.
1045 Wragg, A Free and Regulated Press (n 18) 178.
the relevant discipline they are writing about. They must therefore make do with what O’Neill terms ‘rough and ready versions of the disciplines of truth-seeking.'\textsuperscript{1046} Strict accuracy requirements would therefore create a chilling effect by discouraging reporting that is not based on incontrovertible facts. The press should not be required to be accurate in every case but rather aim for accuracy through regulation that targets processes rather than content. The goals here would be to ensure compliance with professional ethics and epistemic norms such as fact-checking and verification of sources.\textsuperscript{1047} Therefore, the press promoting DNS through, say, misreporting facts about minority groups should only breach accuracy requirements where these norms were not adhered to. This more permissive—but more realistic—understanding of accuracy would have a limited effect on preventing press hate speech. Secondly, even a stringent accuracy provision is only useful where stereotypes are asserted as facts rather than as opinions.\textsuperscript{1048} The columnist who likened migrants to ‘cockroaches’ and ‘norovirus’ in Greer\textsuperscript{1049} and the description of Muslims as a ‘cultural problem’ and linking them to sexual assaults in Elgy\textsuperscript{1050} could not be dealt with under the accuracy clauses because IPSO treated them as opinions.

However, DNS that are not caught by an accuracy clause may be limited based on the second requirement of the deliberative democratic model, which is the need for a rational, diverse and inclusive public discourse.\textsuperscript{1051} There are two reasons to believe that DNS are inimical to pursuing these goals. Firstly, the purpose of deliberation is to arrive at an understanding based on the idea that ‘the authority of the better argument is the only authority’.\textsuperscript{1052} Any recourse to hate speech is contra rationality and, therefore, contra the goals of public discourse.\textsuperscript{1053}

Secondly, DNS are also inimical to a diverse and inclusive public discourse. Recall that hate speech, including DNS, silences groups. It can cause them to withdraw from society and refrain from participating in public discourse. Baker makes the point that

\begin{itemize}
  \item \textsuperscript{1046} Madianou and Pinchevski (n 1038) 24.
  \item \textsuperscript{1047} O’Neill (n 267) 5.
  \item \textsuperscript{1048} Wragg, A Free and Regulated Press (n 18) 193–194.
  \item \textsuperscript{1049} Greer v The Sun [2015] IPSO 02741-15.
  \item \textsuperscript{1050} Elgy v The Sun [2016] IPSO 17562-17.
  \item \textsuperscript{1051} Oster (n 24) 35.
  \item \textsuperscript{1053} Oster (n 24) 30; Post, ‘Racist Speech, Democracy, and the First Amendment’ (n 551) 275.
\end{itemize}
‘if discourse is to be inclusive, then speech that denigrates other participants in the debate is not helpful’. Restricting such speech then serves the purpose of inclusivity. Recall also that because of their rights as speakers, the scope of hate speech restriction that can be imposed against citizens can only go so far without eroding democratic legitimacy. However, the state is not burdened with the same constraints when it comes to the press who have no recourse to these arguments. Oster makes the same observation noting that hate speech:

*does not contribute to a free and non-coercive and non-excluding argumentative public discourse. Therefore, such speech acts can essentially only be based on the individual autonomy rationale: the liberty of people to blow off steam*. Yet… this rationale only marginally applies to the media.

Unlike citizens, the media can be required to refrain from engaging in DNS to allow for a diverse and inclusive public discourse. Consequently, these models are not just consistent with broader regulation of press hate speech. They require it.

b. The liberal pluralist and complex democratic models

The liberal pluralist model also values political participation but has a different view on both its nature and purpose. While the other models focus on citizens forming a consensus and pursuing a common good, the liberal pluralist envisions democracy as a competition of interests. They see society as comprised of people with different conceptions of the good life, and the goal of democracy is to create a compromise between these competing visions. This informs the nature of political participation as deliberation gives way to competition. People with common interests ought to coalesce, mobilise and bargain for their interests against those of other groups. The role of the media is to put forth each group’s interests in the public forum. This model, therefore, values a partisan media engaged more in advocacy than imparting information. Each group must also have media that advocates for their interests,

---

1055 ibid.
1056 Oster (n 24) 227.
1058 ibid 137–138.
1059 ibid 137.
1060 Rowbottom, *Media Law* (n 42) 23; Christians and others (n 1022) 98.
and it therefore also values external pluralism. The model of ‘complex democracy’
advanced by Baker draws on both models.\textsuperscript{1061} It sees democracy as both an attempt
to pursue a common good and as a competition between groups. It envisions a role
for different media fulfilling each of these functions. A partisan and pluralist media that
advocates for each group’s interests and a media that provides accurate information
to the public that informs rational deliberation.\textsuperscript{1062} I will now consider whether
restricting DNS is compatible with a pluralist and partisan media.

It is often argued that pluralism requires segmented markets and diverse ownership,
and this may require market intervention.\textsuperscript{1063} Baker, Rachael Craufurd-Smith and
Damian Tambini have written about the democratic benefits of dispersing ownership
ranging from preventing the concentration of power to viewpoint diversity and have
laid out the policy measures that can achieve this.\textsuperscript{1064} I do not engage with this debate
more generally. My focus here is on how hate speech can hinder pluralism.

Diverse ownership of the media in a market economy may not lead to a different
editorial output. Wragg argues that ‘there could be one owner, ten, a thousand but all
would be business owners…with ambitions of generating profit, and with obligations
to their shareholders’.\textsuperscript{1065} I have already explored the economic incentives that lead
some outlets to engage in sensationalist stories that denigrate groups. I have also set
out that these stories are cheaper to make and more profitable than quality public
interest journalism. Wragg notes that any new owners are also likely to also produce
such content.\textsuperscript{1066} He has a point, as tabloids outperform the serious press in both
circulation and online audiences.\textsuperscript{1067} This suggests that a diverse editorial output may
require more than just diversifying ownership, at least in the context of the commercial
press, and regulation of DNS may have a part to play in this.

\textsuperscript{1061} Baker, Media, Markets, and Democracy (n 958) 143–153.
\textsuperscript{1062} A similar argument is advanced by Bollinger (n 373).
\textsuperscript{1063} Christians and others (n 1022) 98.
\textsuperscript{1064} C Edwin Baker, Media Concentration and Democracy: Why Ownership Matters (Cambridge
University Press 2006); Smith and Tambini (n 250).
\textsuperscript{1065} Wragg, A Free and Regulated Press (n 140) 73.
\textsuperscript{1066} Wragg, A Free and Regulated Press (n 18) 73.
\textsuperscript{1067} ‘Most Popular Websites for News in the UK: Monthly Top 50 Listing’ (n 722); ‘Most Popular
Newspapers in the UK: Print ABC Circulations for April 2022’ (Press Gazette, 23 May 2022)
<https://pressgazette.co.uk/most-popular-newspapers-uk-abc-monthly-circulation-figures/> accessed
15 June 2022.
Such regulation could serve two purposes. Firstly, regulating DNS means that any new owners would not have recourse to this form of sensationalism to generate sales, and this may force them to diversify their output. Secondly, it creates a level playing field by preventing those who do not trade in this form of profitable journalism from being at a competitive disadvantage. An analogy can be drawn between this and pollution. Imagine two companies $A$ and $B$ which are in the same market, and $A$ is dumping its waste in the river, reducing its costs and therefore making it more competitive than $B$. The state has a legitimate basis to step in and regulate such conduct both to prevent harm and to create a level playing field between $A$ and $B$.

One response to this is that the same goals could be achieved through less intrusive market interventions such as subsidising minority-owned media outlets or giving their views more space in public service media.\textsuperscript{1068} Such interventions may not be sufficient as we have seen that DNS subordinate these groups and that the speech of the subordinated is given less consideration by the majority. Further, subsidising content that appeals to a minority audience has a limited effect as the challenge remains as to the prominence and discoverability of such content amongst the general populace.\textsuperscript{1069} If democracy is a competition of various interests, these market interventions may not be sufficient to provide minorities with the ability to reasonably bargain with the majority. Rowbottom is therefore justified when he observes that some restrictions may be imposed on the press because minorities do not have an ‘equivalent partisan media’ and are therefore ‘excluded from the debate’ and prevented from adequately responding to the negative coverage they receive.\textsuperscript{1070} Such regulation is compatible with the liberal pluralist model as Baker notes that it is consistent with intervention when ‘prejudice against discrete and insular minorities’ curtails the process of bargaining and achieving a fair comprise between groups.\textsuperscript{1071} Finally, these interventions are not mutually exclusive, and there is a role for all of them alongside

\textsuperscript{1068} Rowbottom, ‘Media Freedom and Political Debate in the Digital Era’ (n 1004) 497; Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 620.
\textsuperscript{1069} For an overview of the problem of prominence and discoverability see $M^{\circ}$ Trinidad García Leiva, ‘VoD Platforms and Prominence: A European Regulatory Approach’ (2021) 180 Media International Australia 101.
\textsuperscript{1070} Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 628–629.
\textsuperscript{1071} Baker, Media, Markets, and Democracy (n 958) 138.
the regulation of DNS. Such regulation is arguably consistent with and may even be necessary in order to promote pluralism.

The second requirement, partisanship, arguably presents the greatest challenge for such regulation. A partisan media is a fellow participant in political debate rather than a source of information which feeds that debate. Rowbottom notes that this ‘right’ to be partisan is often used by the press to justify publishing discriminatory content, using the example of the Daily Express’s written evidence to the Joint Committee on Human Rights where they stated:

_as a newspaper we are sceptical about the impact of the asylum system on national life and indeed about the alleged benefits of continued large-scale immigration in general. In this we reflect the overwhelming views of our reader... and should not be—constrained by the sort of rules and regulations on coverage which bind, for example, broadcasters during general election campaigns._\(^{1072}\)

This view would appear to be inconsistent with applying to the press the kind of regulation of DNS that apply to broadcasters. It needs to be taken seriously, as Leveson himself endorsed the view that the press had a right to be partisan.\(^{1073}\) There are three responses to this position. The first is that not all the national and regional press are partisan. Some, such as the Times, The Financial Times and The Independent, expressly brand themselves as non-partisan. Even those outlets that clearly identify themselves as having a partisan political outlook contain a mixture of professional reporting and commentary. Rowbottom, therefore, notes that the national press is ‘somewhere between a partisanship and public service model’, suggesting that it may be appropriate to impose on them hate speech restrictions that are broader than those imposed on the public but narrower than those imposed on the public service media.\(^{1074}\) I set out the specifics of such a measure in the next chapter.\(^{1075}\)

Secondly, there is no necessary connection between partisanship and engaging in DNS. It is entirely possible to engage in robust disagreement with immigration or

\(^{1072}\) Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 629.
\(^{1073}\) Leveson (n 2) 78.
\(^{1074}\) Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 626.
\(^{1075}\) Hare (n 60) 626.
asylum policy and various socially controversial issues without essentialising and assigning undesirable traits to groups of people that foment prejudice against them.

Thirdly, even if the press ought to be fully partisan, it does not follow that they ought to be subject to the same legal treatment as the public. Once again, the scope of hate speech restrictions that can be imposed on citizens is limited by their rights as speakers. The only forms of organisational media that similar rationales apply to are publications by extremist associations or political parties. Baker would treat such associations as a type of organisation whose publications, such as newsletters, represent the values of its constituent members. This would also clearly apply to forms of ‘community media’ owned and run by non-professionals as cooperatives, which reflect their cultural identity and values. However, we have seen that these rationales do not apply to profit-oriented associations such as the institutional press, whose speech cannot be assigned to any speaker. Therefore, we can recognise both that the institutional press has a ‘right’ to be partisan and that this right does not include as broad a claim to engage in hate speech as that possessed by the public or these other associations.

These models of democracy set out the different ways in which the media can perform the function of providing information and ideas to the public. I have argued that regulating DNS can be rationalised and is consistent with these contrasting approaches. Such regulation does not undermine the press’s capacity to perform these functions.

5. Providing a platform for another’s hate speech

All these models envision some role for the media as a platform for citizens to hear each other’s views on issues of common concern. Rowbottom puts it well when he notes, ‘by allowing other people to speak, the public become informed about the range of views in society and the intensity with which they are held’. Here the media is facilitating others’ speech, and the paradigmatic examples of these are letters to

---

1076 ibid 617.
1077 Baker, Human Liberty and Freedom of Speech (n 237) 224.
1078 Christians and others (n 1022) 61.
1079 Baker, Media, Markets, and Democracy (n 958) 160; Rowbottom, Media Law (n 42) 23.
1080 Rowbottom, Media Law (n 42) 21.
editors and comment sections on newspaper websites, call-in radio shows, or even the publishing of third-party commentary.\textsuperscript{1081}

The question, therefore, is what kind of standards should be applied when a press outlet publishes an opinion piece by the Taliban,\textsuperscript{1082} interviews members of racist organisations, or is moderating a comment section on their websites.\textsuperscript{1083} These speakers are not subject to the same institutional constraints as journalists, staff opinion writers and syndicated columnists.\textsuperscript{1084} Their speech is more closely connected to the non-instrumentalist speaker-based justifications explored earlier and which require providing strong protection for hate speech. This suggests that when publishing such content, the press ought to be granted the same leeway as the public.

This platforming role seems to require that any restrictions on DNS ought to exempt the press from liability when they are merely disseminating another’s hate speech. However, this exemption should be conditional on the press challenging such views. We have seen that uncritically platforming a hate speaker can amount to endorsing or licensing of a low type, lending them credibility, legitimacy and the necessary epistemic authority for their hate speech to constitute harm. Such a requirement pursues a legitimate aim and is necessary because it blocks the speaker's authority and denies them the felicity conditions required for them to successfully rank and subordinate their targets. It is also proportionate because it does not impose onerous obligations, nor does it prevent the press from fulfilling their platforming function. It also does not impact the rights of the speaker as even if one accepts the arguments advanced by Sunstein and Scanlon that the media ought to be mandated to provide some access rights to the public, there is no case to be made for access to an uncritical platform.\textsuperscript{1085}

Therefore, the core democratic functions assigned to the press in SRT are compatible with restrictions on DNS if certain exemptions are made for the instances where they

\textsuperscript{1081} ibid.
\textsuperscript{1083} Rowbottom, ‘Extreme Speech and the Democratic Functions of the Mass Media’ (n 23) 621.
\textsuperscript{1084} Rowbottom, \textit{Media Law} (n 42) 21.
\textsuperscript{1085} Sunstein (n 394) ch 3; Judith Lichtenberg, \textit{Democracy and the Mass Media} (Cambridge University Press 1990) ch 11.
provide accurate information in their role as public watchdogs or where they critically platform other hate speakers. We have also seen that the watchdog function entails some protection from direct interference by the state and that this suggests that any restrictions ought to be civil in nature and enforced by an independent regulator.

6. Conclusion

This chapter has explored the various democratic functions assigned to the press in SRT literature. It has been noted that in most cases, the watchdog role and imparting information and ideas do not require engaging in hate speech. As such, press hate speech is often low or no-value under a principle of press freedom. Therefore, DNS regulation would not compromise the press’s ability to perform these functions.

However, there are some cases where it is high-value speech. An important aspect of the watchdog role is the protection and insulation of the press from direct interference by the state. This has two implications for such regulation. Firstly, criminal restrictions which go further than the general law should be off the table as they can be used to punish disfavoured publications. Secondly, any regulatory framework ought to insulate the press from direct or indirect state interference. Finally, the press should be exempted from liability where they publish accurate statements of fact that expose wrongdoing, abuse of office or incompetence. A similar exemption should apply to the platforming function, but the latter should be conditional on the press pushing back against the views of the hate speaker.

As such, these restrictions are either compatible with—or can be designed in ways that are compatible with—the various democratic functions assigned to the press by SRT. Restrictions on DNS can be squared with the libertarian account of press freedom preferred by the industry and the SRT theory preferred in the academic literature.

Having set out why such regulation is philosophically consistent with both views, I will in the next chapter consider the practicalities by setting out what regulation that is both effective against DNS and avoids a chilling effect looks like. I will also consider where such regulation would fit under the current regulatory framework.
Chapter 9: The Content and Structure of Press-Specific Hate Speech Regulation

1. Introduction

Thus far, I have argued that broader regulation of press hate speech is compatible with, and is indeed supported by, normative accounts of press freedom. In this chapter, I move from the philosophical to the practical. I reflect on what the content and form of regulation that is effective at proscribing DNS could look like and whether such regulation would be compatible with the positive law.

Section 2 begins by briefly recapping the weaknesses in the current framework that a new system ought to remedy. I then consider what the content of such regulation could include. I draw on approaches taken by other press councils, the Broadcasting Code, and scholarship and propose a specific provision proscribing DNS, setting out its scope and any applicable exemptions. I then demonstrate how it will apply in practice by using it to re-evaluate some decisions by IPSO and IMPRESS. Section 3 considers the structure of such regulation, including whom it should apply to, whether it should be mandatory or voluntary, and the enforcement powers any regulator ought to have. Finally, section 4 evaluates whether these proposals are compatible with Article 10.

2. The content of press-specific hate speech regulation

Before considering what an effective regulatory approach could look like, it is worth briefly recapping why current regulation is ineffective. I have earlier identified four factors. The first is the narrow scope of existing regulation which either fails to regulate such content entirely (IPSO) or is too narrow to capture it (IMPRESS). The second is the lack of independence from regulated publications. This specifically applies to IPSO. This lack of independence manifests in different ways from a bias towards the publisher's version of events even when they are contradicted by the facts and upholding of only a fraction of complaints.

The third is the lack of sufficient enforcement powers. The fourth is the lack of universal coverage. I have noted that some major publications are not part of either regulator,

---

1086 See Chapter 1
and most publications are regulated by the unapproved regulator (IPSO) and therefore fall outside the Royal Charter system entirely. Effective regulation must remedy these four weaknesses. In terms of content, it should cover diffuse hate speech and be broad enough in scope to catch the problematic content (DNS) while preserving press freedom. In terms of structure, the regulator must be independent of the industry and the state, it must have sufficient enforcement powers, and it must cover most of the relevant publications. I will set out regulation that meets these criteria beginning with content and thereafter, its structure.

It is one thing to show that certain speech causes harm. It is wholly another to show that regulation would only catch the harmful speech without encroaching on legitimate journalistic speech. This is often referred to as the ‘chilling effect’ which Brown aptly describes in the following terms:

*The concern is based on the technical difficulties of framing and wording of precise legislation and on the following hypothesis about human psychological and behavior. If law is expressed in a way that is too unclear for a person of average intelligence to reasonably forecast whether his or her speech falls under it, then to avoid the risk of adverse legal consequences he or she may refrain from saying anything remotely controversial, critical, or provocative.*

Any regulation must be effective and precise enough to allow journalists to be able to know with reasonable confidence the kind of content that is caught (and perhaps more importantly, not caught by the provision and is therefore publishable). Jonathan Heawood and Brigit Morris’s survey of hate speech clauses in European press codes and Alexander Brown’s examination of the cluster of hate speech laws, regulations and codes that prohibit negative stereotyping or stigmatization are both useful starting points. Heawood and Morris point out that various press councils in European countries have imposed broader content restrictions relevant to hate speech than

---

1088 Recall that this is hate speech directed against groups as opposed to targeted hate speech which is that which is directed against specific individuals, See chapter 4.
1089 Hare (n 60) 157.
1090 Brown, Hate Speech Law (n 55) 266.
currently apply in Britain. The German, Austrian, and Polish codes all contain provisions that restrict offensive content. The Irish Press Code is the closest match to the British Press Codes. It contains an incitement provision that mirrors the criminal law but goes further by also proscribing material ‘likely to cause grave offence’.

The focus on offence rather than harm is problematic for reasons set out earlier. More importantly, it is a different rationale for regulation than harm prevention which is what this thesis explores. It is also imprecise as though it may cover DNS, it would also catch speech on a plethora of other controversial social, moral, or political issues and therefore chill much journalistic content.

Another potential candidate for a press-specific hate speech clause is to copy the hate speech provisions in the Broadcasting Code which we have seen to be effective at preventing such content from proliferating in the broadcast sector. I have noted that the strengths of the Broadcasting Code are threefold; a broader hate clause that catches moderate expressions, a focus on stereotyping and the reliance on group-specific guidance.

However, it still suffers from some weaknesses. Firstly, it is complex in that it deals with DNS through multiple clauses. A journalist must be satisfied that the content they publish does not breach both the offence provisions (Rule 2.3) and hate speech provisions (Rule 3.2 and 3.3) of the code. This makes it difficult for them to regulate their conduct. Secondly, Rule 2.3 focuses on offence and is therefore problematic for the same reasons as the above-mentioned European press codes. Lastly, Ofcom relies on group-specific guidance in an ad-hoc manner. The fact that Ofcom will use them as interpretive aids is not set out in the code. This creates uncertainty because the broadcasters in Malik Live and the Islam Channel would not have been

---

1092 For an overview of the various provisions see Heawood and Morris (n 37) 33–34.
1093 German Press Code, s 10.
1095 Poland’s Code of Ethics, cl v.
1096 Code of Practice for Newspapers and Periodicals, principle 8.
1097 See Chapter 4
1098 See Chapter 1 for a discussion of the Broadcasting Code provisions.
aware that Ofcom would use the IHRA definition of Antisemitism to determine whether they breached the code. Even if precedent now exists for its use, there is still considerable uncertainty. Will Ofcom rely on other group-specific guidance, such as those defining and giving examples of manifestations of Islamophobia and Antigypsyism developed by the ECRI, EU and APPG, respectively? If they do intend to rely on them, will Broadcasters receive advance notice? If they intend to only use the IHRA definition, are they then creating a ‘hierarchy of hatred’ where hate speech against some groups is treated more seriously than others?1101 There is much that is useful in Ofcom’s approach, but it can be refined to remedy these weaknesses.

a. A specific provision covering Dangerous Negative Stereotypes

I have identified that the principal form of hate speech that appears in the press is the publication of stereotypes about groups. This is the content that any effective regulation ought to target. Brown has written extensively on negative stereotyping and stigmatization as a genus of hate speech.1102 He has also argued that media regulation ought to target negative stereotyping against groups.1103 To achieve this, he recommended the addition of the following sub-clause to clause 12 of the Editors Code:

*The press must also avoid publishing material that is comprised entirely and overwhelmingly of negative stereotypes or stigmatisation of a group identified on the basis of race, colour, religion, sex, gender identity, sexual orientation or any physical or mental illness or disability.*1104

He argues that this would protect groups while being ‘drawn narrowly enough to prevent excessive restriction’ on ‘issues of public interest relating to groups’.1105

Brown is correct in that a general prohibition on publishing any negative stereotype would be too broad. Therefore, he uses ‘entirely and overwhelmingly’ as qualifiers to

1101 ‘Hate Crime Laws: Final Report’ (n 165) 394.
1102 Brown, *Hate Speech Law* (n 55) 22–23.
1104 ibid.
1105 ibid.
limit its scope. The principle is correct, but the approach is not. The issue is not the number of stereotypes used but rather their content. The focus should be on those stereotypes which both assign the kinds of undesirable traits that are likely to foment prejudice against groups and which suggest that these traits define their very nature, what I refer to as DNS.\textsuperscript{1106} We have seen that DNS vary between groups and are informed by specific historical, social, cultural, and political circumstances. This is where group-specific guidance comes in. They should be relied on by a regulator to determine what is or isn’t DNS. The guidance should be incorporated into the code, and the DNS clause should indicate that the regulator will rely on them. This creates certainty as regulated publications need only examine the guidance and avoid publishing the specific stereotypes set out within them to comply with the regulation.

Much of the work is being done by the guidance, and a potential sticking point is which of them to rely on. As I have pointed out, even reliance on the widely accepted IHRA guidance has been controversial. The APPG guidance on Islamophobia, though accepted by many, is contested for what its opponents perceive as a conflation between criticising religious beliefs and racism.\textsuperscript{1107} Further, these represent a sliver of minority groups. The ideal circumstance would be to arrive at guidance pointing out what amounts to a DNS for each of the protected characteristics, but the nature and details of these and how they will be agreed upon are outside the scope of this thesis. Assuming that such guidance can be developed and agreed upon, I draw from Brown’s definition and the Broadcasting Code and suggest a provision along the following lines:

1. The press must not publish material about a group identified based on characteristics such as race, colour, religion, sex, gender reassignment or identity, sexual orientation or any physical or mental illness or disability which:
   a. Contain negative stereotypes that essentialise groups; and
   b. May in the context in which they were used, be reasonably interpreted as encouraging or justifying prejudice against them.\textsuperscript{1108}

\textsuperscript{1106} See discussion in chapter 1.
\textsuperscript{1107} Chris Allen, Reconfiguring Islamophobia (Palgrave Macmillan 2020) 5–7.
\textsuperscript{1108} This is borrowed with some modifications from the voluntary Code of Practice applied by the Australian Broadcast Company (the Australian Public Broadcaster) which in Rule 7.7 provides ‘Avoid the unjustified use of stereotypes or discriminatory content that could reasonably be interpreted as condoning or encouraging prejudice’ see Brown, Hate Speech Law (n 55) 22.
c. *In determining whether the requirements under (a) and (b) are met, the committee shall rely on the group-specific guidance accompanying this code of practice.*

A few points to note about this provision (‘DNS clause’). Firstly, it borrows the protected characteristics from the Standards Code. The question of what characteristics to include is contested. Paolo Cavaliere notes that the characteristics protected by hate speech laws vary considerably between different international and national frameworks.\(^{1109}\) Heinze has criticised these laws for generally being underinclusive.\(^{1110}\) To avoid this, I have included broader categories than those caught under the stirring up offences and the relevant hate speech provisions in international treaties such as the ICCPR and ICERD.\(^{1111}\) It is also non-exhaustive, leaving room for additional characteristics.

Secondly, it is concerned with the problematic effects produced by DNS which as we have seen, are that they *encourage* prejudice by painting others as dangerous or deviant and *justify* and legitimise it by painting them as inferior and therefore deserving of their lower rank in the social hierarchy.\(^{1112}\)

Thirdly, it does not require subjective intent. I have argued that within the frame of SRT theories, press speech is best understood as commercially driven rather than value-disclosing speech. It would therefore make little sense to inquire whether the journalist, editor or owner *intended* to encourage prejudice. Even if this were not the case, it would still make little sense to consider intent. We have seen that research in cognitive and social psychology suggests that the original encoding of stereotypes and their subsequent usage is largely unconscious. Therefore, Timmer is correct when she notes that ‘intent is not a defining characteristic of stereotyping’.\(^{1113}\) This would also follow the practice in other countries where press codes generally do not consider intent to be a relevant factor.\(^{1114}\) It also makes sense procedurally as adjudications of

---

\(^{1109}\) Paolo Cavaliere, ‘Digital Platforms and the Rise of Global Regulation of Hate Speech’ 9–12.

\(^{1110}\) Heinze, ‘Viewpoint Absolutism and Hate Speech’ (n 468) 565–566.

\(^{1111}\) ICCPR, art. 20(2); CERD, art. 4

\(^{1112}\) See chapter 1


\(^{1114}\) This includes Germany, Austria, Belgium, Poland Heawood and Morris (n 37) 33–34.
code breaches are simple and cost-efficient processes based entirely on written complaints and responses by the publisher. Unlike civil and criminal cases which involve considerable cost and administrative burdens, there are no witness testimonies or rules of evidence, and rarely is there any legal representation.\textsuperscript{1115} Therefore, these committees are ill-equipped to reach authoritative conclusions about intent.

In the place of subjective intent, whether something falls within the DNS clause is determined by the guidance and context. The starting point is a presumption that any negative stereotypes contained within group-specific guidance should be treated as reasonably likely to encourage or justify prejudice against groups. Ofcom takes a similar approach treating anything that falls within the IHRA definition as automatically constituting hate speech.\textsuperscript{1116} This is justified in that such guidance ought to be narrowly drafted to include \textit{only} the set of tropes that are used to encourage and justify prejudice against these groups.

However, the guidance should only be the starting point, and the regulator should take the relevant context into account. This includes factors such as the nature of the publication and the readers’ expectations, i.e. those reading the serious press may have different expectations from those reading tabloids. It should also consider whether any DNS are used in a comedic, satirical or polemic manner.\textsuperscript{1117} Once again, this is not a question of subjective intent but rather an objective test, i.e. would the average reader understand this to be a comedic, satirical, or polemic article? Ofcom also determines context by reference to the likely size, composition, and expectations of the audience.\textsuperscript{1118} This targets the harmful content, allows greater predictability, and avoids speculating—and over-reliance—on the publisher’s stated motives.

Thirdly, to be compatible with press freedom, such regulation must contain exemptions for the press (‘press exemptions’) that shield them from liability when publishing DNS is necessary to perform their democratic functions. Earlier, I noted that the watchdog

\textsuperscript{1115} For an explanation of the IPSO and IMPRESS complaints processes see ‘Our Complaints Process’ (n 171); ‘Complaints FAQ | IMPRESS’ <https://www.impress.press/complaints/complaints-faq.html> accessed 8 July 2022.

\textsuperscript{1116} See decisions in Nadim Malik Live and Islam Channel discussed above

\textsuperscript{1117} Oster (n 24) 50.

\textsuperscript{1118} Ofcom Broadcasting Code, Rule 2.3
role requires the press to be exempted when they publish accurate statements of fact that expose wrongdoing, abuse of office or incompetence by those in ‘government and other powerful institutions’. I also noted that they should also be exempted when fulfilling their duty to impart information on matters of public interest requires them to report DNS made by a third party, e.g. a politician, and where they act as a platform for citizens to exchange views. However, for the reasons explored in previous chapters, this latter press exemption should be conditional on the regulated publication pushing back on and distancing themselves from the views expressed by these third parties. We see similar conditions imposed under Rule 3 of the Broadcasting Code where Ofcom considers ‘whether sufficient challenge is provided to the material’.

The final point is that this provision supplements and does not replace those in the IPSO and IMPRESS codes relating to the use of pejorative and prejudicial language against individuals. This does not risk replicating the complexities in the Broadcasting Code as these provisions serve a different purpose from regulating DNS, namely protecting individuals from psychological harm and emotional distress caused by targeted hate speech. Similarly, the prohibitions on inciting hatred in the Standards Code should be retained to deal with cases where publications stir up hatred without resorting to stereotypes, e.g. ‘group x should be removed from this country’. This provision may also become more practically useful if the law is updated in line with the Law Commission’s recommendations to reform the stirring up offences to allow for prosecution even where moderate language is used.

b. The DNS clause in practice

I will now consider how this provision should apply in practice. To do so, I will use it to re-evaluate some prominent decisions by IPSO and IMPRESS relating to antisemitism and Islamophobia, all of which resulted in a finding that the code was not breached. I will rely on the IHRA and APPG definitions and examples of Antisemitism and

---

1119 Hare (n 60) 610.
1120 See Chapter 4
Islamophobia as the relevant guidance and consider whether it would result in different decisions. As I have set out before, there is some consensus around both definitions.

The first case is IMPRESS’s investigation of two of its regulated publications, the online news websites the Canary and Skwawkbox. IMPRESS launched the investigations of its own accord after the publication of a study into antisemitism in the ‘alternative media’ which examined various articles by the two outlets and alleged that they ‘promoted a negative view of Jews’. IMPRESS conducted a brief two-paragraph analysis of the impugned articles concluding that the material used moderate language and therefore did not breach the code. I will now reconsider this case under the DNS clause. One of the articles was an opinion column titled ‘The inconvenient truths that prove it is not anti-Semitic to compare Israel to Nazi Germany’. The article itself set out at length why the state of Israel, its actions and policies are comparable to those of Nazi Germany, and it argued that ‘Gaza is the modern-day version of the Warsaw Ghetto’.

This amounts to antisemitism under the IHRA definition which notes that while it is legitimate to criticise the actions of the state of Israel and its policies, it is not legitimate to draw ‘comparisons of contemporary Israeli policy to that of the Nazis’. The latter is an example of antisemitism. The definition does not elaborate on this, but the reasons have been set out in scholarship. Some refer to this comparison as ‘holocaust inversion’, where the original victims are now portrayed as the perpetrators. Others treat it as a form of holocaust trivialisation as it minimises the crimes of the Nazis. Of particular relevance is that it is suggested that this comparison is an antisemitic stereotype because in ‘post-war society Nazi behaviour has gradually become the

---

1127 Kenneth L Marcus, Jewish Identity and Civil Rights in America (Cambridge University Press 2010) 56.
paradigm for extreme evil’, and by portraying Israel as a uniquely evil state, it suggests that Jews are ‘demonic and uniquely evil’. The essentialising of Jewish identity and associating it with evil is a long-standing antisemitic trope. This is also an undesirable trait that would encourage prejudice against them and amounts to DNS. There is evidence that holocaust inversion has been used for this purpose by extremist groups in Britain. Furthermore, none of the press exemptions apply. The watchdog exemption is not relevant here because this is an opinion piece and does not report any facts which expose wrongdoing or abuse of power, and neither was the publication acting as a platform for another’s views. As such, the articles fall foul of the DNS clause and breach the code.

The second and third cases are IPSO’s decisions in Manji and Elgy relating to claims of Islamophobia. Both decisions were discussed in depth earlier, and I will not restate the facts here. In Manji, the publication made remarks about a specific Muslim individual and assigned collective blame to Muslims more generally by implying that they sympathised with the perpetrator of a terrorist attack. Elgy, on the other hand, implied that Muslims have a propensity for committing sexual offences. Both complaints were dismissed.

I will now reconsider both under the DNS clause. They relate to Muslims, and I will therefore rely on the APPG definition of Islamophobia. Both articles fall within three of the contemporary examples of Islamophobia in the definition, namely:

1. Making mendacious, dehumanizing, demonizing, or stereotypical allegations about Muslims as such, or of Muslims as a collective group such as…the myth of Muslim identity having a unique propensity for terrorism.
2. Accusing Muslims as a group of being responsible for real or imagined wrongdoing committed by a single Muslim person or group of Muslim individuals, or even for acts committed by non-Muslims.

---

1130 See Chapter 1
1131 Manji v The Sun 05935-16 (IPSO).
1132 See discussion in Chapter 1
3. Using the symbols and images associated with classic Islamophobia… to characterize Muslims as being ‘sex groomers’, inherently violent or incapable of living harmoniously in plural societies.\textsuperscript{1133}

\textit{Manji} falls within the first example as it implies Muslims have a unique propensity for terrorism. It also falls within the second example as it assigns collective blame to Muslims. Both also fall within the third example. In \textit{Manji}, the publication implied that Muslims are inherently violent, and in \textit{Elgy}, it implied that they are ‘sex groomers’. Both are also essentialist tropes that present these undesirable traits as inherent to Muslim identity.

The articles can reasonably be interpreted as encouraging or justifying prejudice against Muslims. This is a point borne out by research. Ella Cockbain and Waqas Tuqail have pointed out the Islamophobic stereotypes that Muslims (particularly Muslim men) are fanatics that are prone to violence and committing ‘sexual acts motivated by a patriarchal, misogynistic culture and backward, barbaric religion’ has been used in extremist propaganda in Britain to demonise Muslims.\textsuperscript{1134} Similar to the previous publication, none of the press exemptions apply here, and the DNS clause has therefore been breached.

In previous chapters, we have seen that one of the greatest challenges to effective hate speech regulation—and which gave birth to new racism—is the modification of language to avoid bans. Group-specific guidance isn’t perfect and is less useful where categories are too generic and diverse such as migrants or asylum seekers. The regulator should therefore interpret the clause purposively and consider whether any euphemism is being used to obfuscate the meaning. For example, the ECRI has noted that ‘asylum seeker’ and ‘refugee’ have become synonymous with ‘Muslim’ in European public discourse.\textsuperscript{1135} An example of this is the IPSO decision in \textit{Greer}.\textsuperscript{1136} Here the columnist referred to migrants throughout the article (who she compared to ‘cockroaches’, the ‘norovirus’ and described as a ‘plague of feral humans’) rather than

\textsuperscript{1133} ‘Islamophobia Defined: The Inquiry into a Working Definition of Islamophobia’ (n 183) 56.
\textsuperscript{1134} Ella Cockbain and Waqas Tuqail, ‘Failing Victims, Fuelling Hate: Challenging the Harms of the “Muslim Grooming Gangs” Narrative’ (2020) 61 Race & Class 3, 5, 10–12.
\textsuperscript{1136} See Chapter 1 for a discussion of the case.
Muslims. However, references to North African migrants and ‘sharia stoning’ clearly indicated that the rhetoric was directed at Muslims, and the APPG guidance on Islamophobia should therefore be relied on.

This also points to the importance of regularly updating the guidance, work which is already being done in some cases. For example, the EU’s Handbook for the practical use of the IHRA Working Definition of Antisemitism notes that recent antisemitic tropes use George Soros, a philanthropist of Jewish heritage, as a euphemism to refer to Jews generally.\footnote{Handbook for the Practical Use of the IHRA Working Definition of Antisemitism (n 88) 12.}

Finally, the clause concerns harm, not offence, and publications are free to offend, provoke, critique, and exaggerate.\footnote{Oster (n 24) 50.} Take for example, a controversial polemic in the \textit{Spectator} magazine which suggested that parliamentary elections should be set on a date when ‘Muslims are forbidden to do anything on pain of hell’ as this would ‘deliver at least 40 seats to the Tories’.\footnote{If You Do One Thing This Election, Stop Your Kids Voting | The Spectator’ \url{<https://www.spectator.co.uk/article/if-you-do-one-thing-this-election-stop-your-kids-voting>} accessed 9 July 2022.} This may be offensive, but it does not breach the DNS clause. It certainly stereotyped the voting patterns of British Muslims and was negative in that it essentialised them by treating them as a monolith. However, this trope is not in the APPG definition; therefore, it does not constitute an undesirable trait that encourages or justifies prejudice against Muslims. There is also no suggestion that Islamophobia manifests this way in Britain, historically or contemporarily.

The DNS clause is therefore effective at targeting harmful content while leaving the editorial freedom to publish offensive content intact. It also creates certainty by clearly specifying the content to which it applies, allowing publications to regulate their behaviour and avoiding a chilling effect. Finally, it serves an important educative and preventative function as editors and journalists are made aware of the various ways prejudice can be stirred up against minority groups whose cultures and customs they may be unfamiliar with.
3. The structure of press-specific hate speech regulation

This DNS clause would not serve much purpose if a regulator (or the regulators) did not cover the whole (or most) of the industry or if they did not have sufficient enforcement powers.1140 Here I consider different regulatory structures setting out those which are the most suitable for achieving these goals while preserving the independence from the state required by the watchdog function. Any system of regulation must answer three questions. Who should it apply to? What form should it take? And what enforcement powers should the regulator have?

a. Who should it apply to?

It would certainly apply to the press and not broadcasters who remain regulated by Ofcom. In Chapter 1, I noted that while the functional definition of the press is preferable when determining whether a person/institution carrying out certain activities ought to benefit from special privileges, a regulatory framework should specify the institution that it will apply to. Here I look at two possible approaches to determine what institutions should be covered by such a framework.1141

The first is simply to cover entities that are already part of the existing regulators, i.e. existing members of IPSO and IMPRESS. The benefit of this is that it will cover most of the national and regional press, but it does not achieve the universal coverage sought by Leveson as three publications with significant circulation are not a member of either regulator (the Guardian, the Independent and the Financial Times). However, an aspect of the approach taken by IMPRESS and IPSO which should be adopted is the holistic nature of the regulator’s ambit which covers the regulated publication’s offline and online output, save that which falls within the remit of Ofcom.1142

The second approach is to distinguish between commercial and amateur entities, with regulation only covering the former.1143 This better fits with the normative arguments advanced in previous chapters. The speaker-based justifications explored in chapter

---

1140 Leveson (n 2) 1758, 1771,1782,1794.
1141 These are inspired by Rowbottom’s approaches to determine what kind of new media ought to be subject to similar regulation as the traditional media Rowbottom, ‘Media Freedom and Political Debate in the Digital Era’ (n 1004) 508–509.
1142 See Chapter 3
1143 Rowbottom, ‘Media Freedom and Political Debate in the Digital Era’ (n 1004) 509.
7 apply to lone bloggers and citizen journalists. We have seen that in these cases, the demands of political legitimacy limit the scope of hate speech restrictions that can legitimately be applied to such speakers, i.e. only those which target overt hate speech while leaving intact alternative avenues for the speakers to communicate their views and values. Similar arguments apply to community and political party media which should also be excluded from a regime restricting DNS.

This may be problematic from a harm prevention standpoint. A principal issue with the media is that their communicative power gives them the capacity to cause greater harm than most lone hate speakers. Some citizen journalists and ‘star bloggers’ may also have a wide reach, and their hate speech may therefore cause similar direct and indirect harm as the institutional press. They may also develop parasocial relationships with their followers and have the necessary practical, and epistemic authority for their speech acts to constitute harm.

However, their influence should not be overstated, nor should it be claimed that it is equal to or greater than the national or regional press. Firstly, the institutional press does much of the original reporting, and bloggers depend on them for content, with the latter primarily providing commentary and opinions based on original reporting by the press. The editorial choices of the institutional press and what they decide to report (or not report on) set the agenda and frame the terms of the debate. Secondly, as noted earlier, the institutional press are repeat players with the staff numbers and expertise to publish more content more often. Rowbottom concluded that even in an era of increased user-generated content, the capacity of the mass media and media elites to influence public debate and opinion is unlikely to change. This view expressed in 2006 remains relevant to date, even with the profound changes in digital technologies and the ascendance of social media platforms. In the UK, institutional players still hold a monopoly on news sources. 15 out of the top 20 used news sources are forms of institutional media. Most of the other sources are social media

---

1144 ibid 503.
1146 Rowbottom, ‘Media Freedom and Political Debate in the Digital Era’ (n 1004) 507.
1147 ibid.
1148 ‘News Consumption in the UK: 2022’ (n 50) 19.
platforms such as Meta which increasingly rely on and amplify ‘trusted news sources’ and original reporting from institutional actors.\textsuperscript{1149}

Furthermore, once a ‘star-blogger’ becomes influential enough, they will likely adopt a corporate structure that resembles the institutional press. The greater the demand for content, the less likely they will produce it themselves. They may hire additional staff or even start to host content by other bloggers and play a gatekeeping role.\textsuperscript{1150} The star blogger then begins to resemble the traditional media.\textsuperscript{1151} At the same time, the speaker-based arguments for free speech become more ill-fitting. An example of this is \textit{Guido Fawkes} which started as a political blogging website run by a single person but now has considerable reach, influence, and multiple contributors.\textsuperscript{1152} In some sense, a focus on commercial entities will select for forms of new media that have the success and communicative power necessary to cause these various harms. However, this cannot be the sole criterion. It would be unjustifiable to apply a regime to a blogger with a limited reach who is trading as a single shareholder/director company but not to an influential one running a successful unincorporated business. We must refine it further.

Comprehensive criteria can be found in existing and proposed legislation. The first is the Crime and Courts Act which defines the press entities that ought to be subject to the Royal Charter System. The second is the proposed Online Safety Bill which defines the press in order to identify the entities which are excluded from the duties in the proposed legislation. Both contain similar criteria referring to those outlets which publish ‘news-related material in the course of business (whether or not carried on with a view to profit)’\textsuperscript{1153}, which is ‘written by different authors’\textsuperscript{1154} and is ‘subject to

\begin{itemize}
\item \textsuperscript{1149} See generally James Meese and Edward Hurcombe, ‘Facebook, News Media and Platform Dependency: The Institutional Impacts of News Distribution on Social Platforms’ (2021) 23 New Media \& Society 2367;
\item \textsuperscript{1150} Rowbottom, ‘Media Freedom and Political Debate in the Digital Era’ (n 1004) 502.
\item \textsuperscript{1151} ibid 503.
\item \textsuperscript{1152} Declan McDowell-Naylor, Stephen Cushion and Richard Thomas, ‘The Role of Alternative Online Political Media in the 2019 General Election’ in Dominic Wring, Roger Mortimore and Simon Atkinson (eds), \textit{Political Communication in Britain: Campaigning, Media and Polling in the 2019 General Election} (Springer International Publishing 2022) 151–152.
\item \textsuperscript{1153} Crime and Courts Act 2013, s 41(1); Online Safety Bill, Clause 50 (2) (b).
\item \textsuperscript{1154} Crime and Courts Act 2013, s 41(1) (a); Online Safety Bill, Clause 50 (2) (a) (i).
\end{itemize}
editorial control’. News-related material includes news and current affairs, opinion, and gossip about celebrities and public figures.

The Crime and Court Act contains an important exclusion for a ‘micro business’ that is either a ‘multi-author blog’ or which ‘published news incidental to their business’. This is relevant here as micro businesses are defined in law as those having less than ten staff members and who have a turnover of less than £2 million. This could be used to exclude those smaller entities (at least from the mandatory model discussed below) that do not have the reach or communicative power required to produce the various harms set out in this thesis. It also targets those entities that are more likely to be profiting from the publication of DNS. Finally, the publisher should have a registered office in the UK and have legal responsibility for material published within it. The democratic functions assigned to the press determine the appropriate scope of regulation, and as Rowbottom points out, the needs of British democracy should not be imposed on publications based abroad that target a non-UK audience. The Council of Europe’s Recommendation on a ‘new notion of media’ is also useful in picking out the types of new media entities that should be regulated. It includes six criteria which significantly overlap with the ones set out above, including whether the entities in question hold themselves out as media, disseminate media content, have editorial policies and processes, follow professional standards, seek to reach a ‘large number of people’ and meet public expectations.

Definitions are always plagued with concerns about under or over-inclusiveness, and that is no different here. For example, there is nothing magical about the £2 million threshold, and neither is it set in stone. It is not as if the speaker-based arguments for free speech cease to apply once turnover crosses this threshold. But there is no avoiding the need to have definitive criteria with a clear cut-off that allows for certainty. This will inevitably involve a degree of arbitrariness, but perfection should not be the enemy of good. The above criteria (‘relevant criteria’) do cover the institutional press

1155 Crime and Courts Act 2013, s 41(1) (b); Online Safety Bill, Clause 50 (2) (a)(ii).
1156 Online Safety Bill, Clause 50 (5).
1157 Crime and Courts Act 2013, sch. 15, para 8
1158 Small Business, Enterprise and Employment Act 2015, s 33
1159 Online Safety Bill, cl 50(2) (e) (f).
1160 Rowbottom, ‘Media Freedom and Political Debate in the Digital Era’ (n 1004) 509.
1161 Recommendation CM/Rec (2011)7 of the Committee of Ministers to member states on a new notion of media
and many of the new media players that closely resemble them. It benefits from being part of existing law (or legislative proposals) and will likely meet less resistance. It is also far more specific and targeted than comparable definitions in countries such as Denmark, Germany, and Sweden. It also excludes many of those who are more likely to have recourse to the speaker-based arguments for free speech that would militate against such regulation in the first place. Those not captured can voluntarily sign up and can even be incentivised to do so. I consider these incentives below.

b. What form should it take?

While the broadcast media in Europe is mostly subject to statutory regulation, there is much greater variation in the regulatory models applied to the print and online press. Wragg examines nine European countries finding significant differences between them though pointing to the primacy of self-regulation. Perhaps the most expansive study in this area is Lara Fielden’s comparative examination of international press councils, where she sets out a taxonomy of three models found in liberal democracies; voluntary self-regulation, voluntary independent regulation with statutory incentives and co-regulation. I will not consider the first model of voluntary self-regulation as this is what applied in Britain before the introduction of the Royal Charter system. I have already pointed to the deficiencies that make this system undesirable, and I will not repeat that here. Rather, I focus on the other two models which present more promising solutions, and these models are now the preferred approach in Europe. I will examine each setting out their respective advantages and disadvantages. Before doing so, it is important to note that a DNS clause is only a part of a comprehensive code of conduct, and the rest of the obligations will have to be fleshed out. What I hope to do here is sketch out the contours of an appropriate framework where not just a DNS clause but any other—justified—standards can be included.

1162 Wragg, A Free and Regulated Press (n 18) 257.
1163 Ibid 54.
I. Co-regulation

Fielden defines co-regulation as a system established by statute that combines elements of mandatory regulation with self-regulation.\textsuperscript{1165} Such a system is in force in Denmark. Here an independent press council is established by statute.\textsuperscript{1166} Membership is mandatory as all printed publications are automatically included.\textsuperscript{1167} Despite these requirements, this model is co-regulation rather than statutory regulation. This is because not all publications are covered, and online-only outlets must opt into the system.\textsuperscript{1168} These outlets are incentivised to join to enjoy special privileges such as protection of their sources and research, access to information as well as special rights of access to the courts.\textsuperscript{1169} The Press Council has several members from the media, and it has complete discretion to determine the contents of the code, with the only requirement being to ‘provide for sound press ethics’.\textsuperscript{1170} This gives the regulator almost unfettered discretion in deciding its contents.\textsuperscript{1171} This is in stark contrast to Ofcom, which is under a statutory duty to include certain content standards in the broadcasting code.\textsuperscript{1172}

From the perspective of regulating DNS, there are several advantages to this model. Firstly, it achieves Leveson’s goal of near-universal coverage of the industry. The automatic membership could be based on the criteria set out above. Secondly, it achieves Leveson’s vision of a single regulator covering the whole industry.\textsuperscript{1173} A single regulator applying the same DNS clause allows for the building of institutional knowledge, expertise, and consistency in adjudicating such content. There are also concerns that a multiplicity of regulators could lead to perverse incentives as the press would join those regulators with the laxest standards. As funding is—at least in part— informed by the size of the membership, other regulators would, in turn, be incentivised to lower their standards to compete for members, creating a race to the bottom.

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{1165} Fielden (n 451) 39.
\item \textsuperscript{1166} Media Liability Act (Denmark) 1991.
\item \textsuperscript{1167} Ibid, s 1.
\item \textsuperscript{1168} Fielden (n 451) 28.
\item \textsuperscript{1169} Ibid 53.
\item \textsuperscript{1170} Ibid 28–29.
\item \textsuperscript{1171} Ibid 52.
\item \textsuperscript{1172} Communications Act 2003, s. 319.
\item \textsuperscript{1173} Leveson (n 2) 1758.
\end{itemize}
\end{footnotesize}
An aspect of the Danish model that should not be borrowed is the composition of the committees that adjudicate complaints. These should be professionals who have both requisite experience and independence from industry and the state. As an example, members of Ofcom’s Content Board are required to avoid employment with any media organisation currently being regulated by Ofcom.

While the specifics of the code should be determined by the regulator, the statute should impose on them a duty to set rules to protect groups from discriminatory coverage. This could be framed in terms of Leveson’s recommendation number 38 which required the regulator be equipped with ‘the power to intervene in cases of allegedly discriminatory reporting’.\textsuperscript{1174} I have argued that one of the ways of achieving this is through the DNS clause set out above, but it should be up to the regulator to achieve this goal through this or any alternative means that is equally effective.

Such mandatory models are often critiqued by the press on three counts. The first is a slippery-slope argument that sees any involvement of the state in press regulation as laying the foundation for future censorship.\textsuperscript{1175} The fear is that even if regulation of DNS is legitimate, allowing this degree of state involvement sets a precedent, as it is much easier to amend existing law to cover more content than it is to introduce new legislation.\textsuperscript{1176} There are certainly ways of minimising this risk. For example, the statute could be entrenched such that any amendments would require more than a simple majority of parliament. This does not, of course, address the concern that governments with big majorities can always legislate in ways that undermine press freedom, but that is the reality of the British Constitution, mandatory scheme or not.

The second critique is that for a mandatory scheme to work, the ultimate sanction for a breach would involve either criminal proceedings or banning an outlet from publishing.\textsuperscript{1177} The fear of criminal sanctions has some grounding. In Denmark, the failure to comply with the regulators’ directions, e.g. failing to publish a correction, could result in a fine or up to a four-month prison term.\textsuperscript{1178} The other fear (banning an outlet) is less justified. There is no precedent for this in the countries reviewed by

\textsuperscript{1174} ibid 1808.
\textsuperscript{1175} Wragg, A Free and Regulated Press (n 18) 267.
\textsuperscript{1176} Fielden (n 451) 39.
\textsuperscript{1177} Wragg, A Free and Regulated Press (n 18) 272.
\textsuperscript{1178} Fielden (n 451) 53.
Fielden. Wragg also points out its implausibility. Unlike a broadcaster, the press cannot be denied access to a spectrum, and therefore, they would have to ban either (or a combination of) the outlet, the journalist, the editor or the owner, and this seems far-fetched. Tambini is right to point out that it is incumbent upon those presenting slippery-slope arguments to provide an ‘explanation of lubrication and gravity’, and the press have not done so here.

Therefore, co-regulation is, therefore robust enough to withstand these critiques. This model’s biggest weakness is not its justifiability but rather realpolitik. A mandatory scheme stands little chance of being enacted into law. Wragg points out that no government would countenance enacting it for fear of getting on the wrong side of the press and ‘jeopardising the prospect of re-election’. This scepticism is certainly justified by decades of failed inquiries and ‘last chance saloons’, reflected most recently by the government’s cancellation of the second stage of the Leveson inquiry due to pressure from the press.

II. Voluntary independent regulation with incentives

The second system is voluntary independent regulation with incentives (‘voluntary regulation’). Under this arrangement, either a regulator or a system to recognise and approve regulators is set up by statute. Publications are not mandated to join the regulator or any approved regulator, and membership is instead incentivised. An example of such a system is in place in Ireland where statute recognises the Press Council and sets out its minimum requirements. Publications that sign up to the regulator and who have a track record of compliance with the code are then able to, more easily, rely on a defence of ‘fair and reasonable publication’ in defamation proceedings.

---

1179 Wragg, A Free and Regulated Press (n 18) 273.
1180 Tambini (n 22) 25.
1181 Wragg, A Free and Regulated Press (n 18) 266.
1182 Leveson (n 2) 1757.
1184 Defamation Act (Ireland) 2009, s. 44.
1185 Fielden (n 451) 47–48. Though as Wragg points out, the defence has had a limited impact in practice Wragg, A Free and Regulated Press (n 18) 58.
This system is also in force in the UK, though underpinned by the Royal Charter rather than by statute. The Charter also establishes a system to recognise regulators rather than a single regulator. To qualify for recognition, a press regulator should have both a complaints process (to deal with breaches of the code) and an arbitration service to provide claimants with an inexpensive avenue to bring civil claims against the publisher in areas such as defamation and misuse of private information. Members of approved regulators were to be protected from exemplary damages as well as benefit from favourable cost-shifting provisions in any civil claims brought against them in the courts. Those who were not a member of an approved regulator could be subject to exemplary damages and had unfavourable cost-shifting provisions that would see them liable for the claimant's costs even if the claim was dismissed. The cost-shifting provisions which are the most consequential of these incentives, have never come into force and are set to be repealed largely due to successful lobbying by the press.

DNS regulation can be fit into the current UK system with only a few modifications to the existing framework. Currently, the Royal Charter empowers the PRP to recognise regulators who satisfy the recognition criteria contained in the charter. The criteria which ensure any approved regulators are independent of the industry are robust and ought to be maintained. The requirement to have in place a framework to effectively regulate DNS could be included as part of the recognition criteria. Notably, this will not require wholesale changes. Leveson recommendation 38, which I have argued requires such measures, is already included in the Royal Charter. However, it is currently only a desirable rather than an essential requirement for recognition. I would therefore amend the Charter (or include it in a statute, which I argue below

1186 Leveson (n 2) 1781.
1187 Crime and Courts Act 2013, s. 34 (2), 40 (2).
1188 Ibid.
1191 Royal Charter on Self-Regulation of the Press, schedule 2, clause 4.
should replace the Charter) to turn this recommendation into a condition for recognition.

c. Discretionary benefits and burdens: towards a robust incentive structure

The discretionary benefits are a key part of the voluntary model. These incentives are crucial to the success of this model are they are the only means by which to secure membership, and they must therefore be powerful enough to realise this goal. Here I look at three incentives. Preferential VAT treatment for audio-only content produced by press outlets, measures that aim to increase the press’s negotiating power against online platforms and news aggregators, as well as exemptions from the upcoming Online Safety Bill and similar legislation.

It is useful to distinguish between incentives and burdens. Incentives reward regulated publications by granting them preferential treatment relative to other speakers under the law. Press outlets are therefore treated the same as any other publisher if they choose not to join an approved regulator. This is pertinent as conceptually, incentives would then not count as a ‘restriction’. Burdens, on the other hand, punish the press by imposing sanctions on unregulated outlets. These burdens are coercive in that the press is treated worse than other publishers and the public if they choose not to sign up to the regulator. These burdens are restrictions. Though I have argued that coercive restrictions on DNS are normatively justified, there is significant room for debate and disagreement. There is, however, much less room for principled opposition to measures that do not restrict the press but rather provide them discretionary benefits as a reward for taking on duties to behave ethically and responsibly.

Some incentives, such as a kitemark to indicate that a publication is a member of an approved regulator, are very straightforward.1192 However, while useful for smaller publications and new entrants, they have limited utility in incentivising larger publications with established readerships to join the regulator.1193 Economic incentives are more powerful. We have seen that the publication of DNS is largely attributable to a profit motive, and the externalities such content produces are comparable to

1192 Fielden (n 451) 100.
1193 Leveson (n 2) 1659; Wragg, A Free and Regulated Press (n 18) 260.
pollution. Economic incentives that financially reward publications which agree not to pollute and burdens which make it less profitable to do so can dissuade this form of profitable but unethical journalism.

One economic incentive is an adjustment to VAT treatment. VAT is a general tax that applies to the supply of goods and services. Print newspapers have been zero-rated since 1972, and subscriptions to online publications since 2020. However, the zero rating for online publications does not extend to audio content such as e-audio books and subscriptions to online radio services and podcasts which are still charged at the standard rate. I have in chapter 3 noted that press outlets are increasingly producing audio-visual material alongside their articles, and this material is often unlikely to be within the scope of Ofcom regulation. They are also now increasingly offering audio-only podcasts either as part of an article or as standalone services. Podcasts are entirely out of the scope of Ofcom regulation even if they are offered as standalone services as they are not audio-visual material, and neither do they require a radio broadcast license. Ofcom podcast consumption surveys also indicate that people are now increasingly consuming such content with these listening figures likely to increase. To incentivise membership, subscriptions to any audio-only services, such as podcasts offered by regulated publications, could be zero-rated rather than standard-rated for purposes of VAT.

Notably, this would qualify as an incentive rather than a burden, as publications that choose not to join the regulator would simply be subjected to the same VAT treatment as any other publisher. This does not implicate freedom of expression or press freedom, as neither principle includes a right for the press to be subject to a more beneficial tax regime than members of the public. Rather, the objections to differential VAT treatment raised during the Leveson inquiry (and subsequently) are that it would eat into the profits of unregulated publications. But as Leveson noted, this is a feature, not a bug. The purpose of economic incentives is to make

---

1195 As an example, The Times, The Guardian, and The Mail brands of newspapers etc all have podcast services.
1196 ‘News Consumption in the UK: 2022’ (n 50) 15.
1197 Barendt, Freedom of Speech (n 6) 429.
1198 Leveson (n 2) 1660; Wragg, A Free and Regulated Press (n 18) 260.
1199 Leveson (n 2) 1660.
regulated publications financially better off than unregulated ones. Otherwise, there would be little purpose in joining a regulator. The biggest concern—and why this proposal was ultimately shelved—was that EU law on VAT and state aid rules limited a member state’s discretion to apply differential rates to similar goods. It is arguable then that this barrier no longer applies after the UK’s exit from the EU.

Another potential set of economic incentives are those which aim at increasing the negotiating power of the press against online platforms and news aggregators for the use of their content. Changing consumption patterns have led to a decline in print and the advertising revenue that it generates. The press cannot compete with the tech giants such as Google and Meta in online advertising as the latter are more attractive to advertisers because they hold more granular data about consumers for targeted advertising. At the same time, the press relies on search engines and social media platforms for content discovery and to direct traffic to their websites and apps. This results in a significant power imbalance between the press and these platforms in negotiations around payment for the use of the former’s content. Measures that seek to remedy this include the EU ‘press publishers right’ which gives publishers a ‘neighbouring right’ to claim an ‘appropriate share of revenue’ from the use of their publications by these platforms. An alternative solution is Australia’s ‘News Media Bargaining Code’ which seeks to compel platforms to pay for news content, and similar proposals are now at a preliminary stage in the UK. The Australian solution’s strength is its multi-layered approach. Firstly, it imposes transparency obligations which mandate the platforms to provide information on engagement figures which allows the press to arrive at an accurate valuation of their content. Secondly, it

---

1200 ibid.
1202 ibid 34.
1203 ibid 13,22.
1207 ibid 31.
requires platforms to enter into payment negotiations with the press. The press’s bargaining power is enhanced through a mandatory arbitration scheme where disputes in valuation are settled by the press outlet and the platform each tabling an offer with the tribunal having the final decision on which valuation to accept.\textsuperscript{1208} The power imbalance in the current system often means that the press is pressured into accepting undervalued offers for their content. This solution attempts to remedy this by levelling the playing field and putting the press on a more equal footing with these platforms and news aggregators. Of note is that to benefit from this framework, the press outlet must be a corporation, have a minimum amount of turnover and, most importantly, must be under a regulatory regime.\textsuperscript{1209} Similarly, the UK government could extend these benefits only to members of approved regulators.

The final incentive I will consider relates to the press exemption in the pending Online Safety Bill. The bill imposes duties on platforms to act on legal but harmful content, a category that will be set out in secondary legislation.\textsuperscript{1210} They are not required to remove such content but must clearly set out how they will deal with it in their terms of service, and they must enforce the terms consistently or face enforcement action.\textsuperscript{1211} This is relevant because platforms already restrict what Daphne Keller terms ‘lawful but awful’ content because it alienates users and advertisers.\textsuperscript{1212} Many have restrictions in their terms of service that cover DNS which these proposals will effectively encourage them to enforce more rigorously.\textsuperscript{1213} Meta prohibits the use of ‘harmful stereotypes’ which have ‘historically been used to attack, intimidate, or exclude specific groups’.\textsuperscript{1214} Twitter prohibits ‘targeting individuals and groups with content intended to incite fear or spread fearful stereotypes about a protected

\begin{footnotesize}

\textsuperscript{1208} Lee and Molitorisz (n 1205) 42–43.
\textsuperscript{1209} ibid 40.
\textsuperscript{1210} Online Safety Bill, cl 13, ch 6
\textsuperscript{1211} ibid
\textsuperscript{1213} Cavaliere (n 1109) 14–16.
\end{footnotesize}
category’. YouTube limits the ‘use of racial, religious or other slurs and stereotypes that incite or promote hatred’.

The press benefit from beneficial treatment under the Bill in two ways. Firstly, content on their own websites as well as their articles that are shared on these platforms are outside the scope of the Bill. Platforms are also under a duty to provide protection to ‘journalistic content’. The government should instead treat the exemption and additional protection to journalistic content as discretionary benefits extended only to press outlets that are part of an approved regulator.

At first glance, the press may not seem to gain much from this. After all, the effect of this would be to substitute DNS regulation by platforms with DNS regulation by the approved regulators. However, there are strong incentives for the press to choose the latter. Firstly, platform regulation of this content is underdeveloped as it does not include group-specific guidance. Secondly, the terms of service do not contain exemptions for the press for the watchdog and platforming functions. Thirdly, platforms are also moderating at scale and will be forced to rely on blunt tools, including algorithms, to detect such content. This has the potential to over-censor and create a chilling effect. Therefore, if press outlets will inevitably be subject to such regulation, it is in their interest to ensure that it is being applied by an independent, specialised regulatory body interpreting a contextualised and narrowly drawn DNS clause that targets only the content that causes harm.

d. The need for statutory underpinning

Co-regulation would certainly require a statute to establish the regulator and its powers as well as mandate membership. The question is whether this is also necessary for voluntary regulation. Barendt suggests that legislation is not necessary for the independence of a regulator or the establishment of a recognition body, as this could

---

1217 Online Safety Bill, cl 39(8); ‘Joint Committee on the Draft Online Safety Bill’ (House of Lords 2021) HC 609 para 286. Para 286
1218 Online Safety Bill, cl 14.
easily be achieved through the Royal Charter. The charter arrangement benefits from a status-quo advantage, and refinements to it would encounter less resistance than the introduction of a statute. The press did indeed successfully push for the adoption of a charter scheme in the first place as an alternative to Leveson’s proposals for statutory underpinning, as they viewed the former as less intrusive on press freedom.

However, the charter system suffers from a glaring shortcoming which makes it possible for the state to directly interfere with the independence of the regulatory body. The charter is said to be ‘entrenched’ in order to protect it from such interference. Legislation can be amended or repealed by a simple majority of either a single or both houses of the UK parliament. However, certain legislative provisions can be framed in terms that impose more onerous requirements, in which case such a law is deemed to be entrenched. An example is the Scotland Act 2016 which sets out that the Scottish Parliament and Government cannot be abolished without a referendum of the people of Scotland. The independence of the Press Recognition Panel is also said to be ‘entrenched’ through the charter system. Clause 10 of the Royal Charter requires that any amendments to it require a two-thirds majority of both Houses of Parliament and that of the Scottish Parliament as well as the unanimous agreement of the board of the Press Recognition Panel. This is then supplemented by statute through section 96 of the Enterprise and Regulatory Reform Act, 2013 which provides:

Where a body is established by Royal Charter after 1 March 2013 with functions relating to the carrying on of an industry, no recommendation may be made to Her Majesty in Council to amend the body’s Charter or dissolve the body unless any requirements included in the Charter on the date it is granted for Parliament to approve the amendment or dissolution have been met.

In order to prevent even the appearance of statutory underpinning, the entrenched provisions are only provided in the charter. The result is that the statute only prevents

---

1220 Heawood (n 1190) 132–133.
1222 Scotland Act 2016, s. 1 (3). There is an unsettled debate about whether entrenchment is compatible with the UK Constitution see generally Blick (n 1221).
1223 Royal Charter on Self-Regulation of the Press, cl 10
the government from amending the charter in a specific way, i.e. by recommending to
the Queen that the Royal Charter be amended. However, section 96 itself is not
entrenched, meaning nothing prevents the government from simply repealing or
amending it through legislation passed by a simple majority and revoking the Royal
Charter itself or reconstituting the Press Recognition Panel and filling it with
government appointees. The result is a sort of rhetorical or performative entrenchment
which is entrenchment in name only. Rather, the recognition panel’s or the approved
regulator’s independence can only be shielded from interference by including these
requirements (two-thirds majority of both Houses etc.) in a statute.

Moreover, even if the charter is sufficient to establish the regulator, a statute is still
necessary for at least some of the incentives set out above, which require amending
existing or upcoming legislation. Finally, several large publications that oppose any
statutory underpinning in the UK subject themselves to the voluntary Irish system,
which as noted earlier, is underpinned by statute and their objections to a statutory
framework are therefore questionable in principle and, in some cases, appear to be
disingenuous.

e. The enforcement powers

This section considers the different kinds of enforcement powers used by press
regulators to sanction breaches of the code and set out what additional powers—or
modifications to existing ones—are necessary to better ensure compliance.

A common enforcement power seen in various countries such as Finland, Sweden
and Denmark is the ability to direct the regulated outlet to publish the fact of an adverse
finding and/or a correction. Some regulators, including both IPSO and IMPRESS, have
set requirements on where this content will appear on the print and online versions of
the respective outlets. Corrections are largely reserved for inaccuracies, and we
have seen that accuracy clauses are limited when DNS are expressed—as they often
are—as opinions rather than facts. However, IMPRESS has used it in a hate speech

---

1226 Heawood (n 1190) 131.
1227 IMPRESS Standards Code 2017, cl. 1.2; IPSO Editors Code, cl 1(ii).
case where a correction was understood to mean amending the contents of the publication to remove the sections that breached the hate speech clause. A similar power can be used to direct the removal of content that breaches the DNS clause.

Tom O’Malley and Clive Soley have praised the use of such measures arguing that the fear of publishing numerous corrections would incentivise outlets to comply with the code. The view is that publishers will be afraid of ceding valuable space in print and that the appearance of prominent notifications on the homepages of their websites which indicate that they have failed to meet regulatory standards damages their brands. This has proven to be effective in the Scandinavian press, where the fear of reputational damage caused by the mere upholding of a complaint has been sufficient to ensure compliance. However, Wragg points to the absence of a similar culture in the British press, where antipathy is a more common response—particularly by large publishers—to adverse findings.

Therefore, this remedy should be bolstered by additional requirements such as a right to reply. The purpose of such a right is to give the groups which are subject of hate speech a platform and opportunity to respond to the negative coverage. I have earlier pointed out that counter-speech has limited utility when the group that is speaking back is doing so from a position of subordination as their speech is taken less seriously. It is certainly questionable whether a group which is regularly vilified by a publication would receive a fair hearing from its readership. However, this does not mean that counter-speech has no effect. We have seen that the mediating factors that limit the effect of media stereotypes on outgroup prejudice include intragroup contact and knowledge. Giving these groups access to a prominent platform both creates intragroup contact and provides them with an avenue to educate and correct any misconceptions the readership may have about their identities, cultures, customs and/or beliefs. This then equips the readers with the tools to counter and reject stereotypical depictions by the media. It also alleviates the concerns about these groups lacking an equivalent powerful partisan media to respond to these attacks.

1228 5 Pillars (n 201) 18-19.
1229 O’Malley and Soley (n 139) 189.
1230 Fielden (n 451) 74.
1231 Wragg, A Free and Regulated Press (n 18) 59–60.
1232 Hare (n 60) 628–629.
There is a potential sticking point in determining who should get to reply. One of Leveson’s recommendations on discriminatory reporting—and which is included as a recognition criterion in the Royal Charter—was that regulators should have the capacity to hear complaints from ‘representative groups’. IPSO has also heard complaints from a representative group in the Trans Media Watch case. Similarly, IMPRESS accepts complaints from a ‘charity or non-government organisation where that organisation represents a group affected by a potential breach of the Code’. Comparable requirements can be included in a provision containing a right to reply. The representative groups should have some expertise on the matters at issue and have some standing within the communities they claim to represent. Finally, these groups are not a monolith and have many diverging voices within them. A right to reply should not commit the very same essentialising sin as the publication which breached the clause and should reflect diverse views.

The principal objection to such a provision is that it is draconian as it infringes the publisher’s freedom of expression by forcing them to ‘say something against their will’. However, this objection would only make sense if in publishing the journalist, editor, or owner was self-expressing, but we have seen that press freedom is more plausibly viewed as serving audience interest through fulfilling various democratic functions. A right of reply would certainly serve one of these functions by increasing the plurality and diversity of ideas and views the audience is exposed to.

Another objection is that it creates a chilling effect as the press would stop publishing anything controversial that may trigger a requirement to publish a reply. This is even less persuasive because this right is not available on request but rather a discretionary sanction imposed by the regulator when the terms of the code are breached. Finally, this remedy is already recognised in some instruments and is already applied by IPSO and IMPRESS with respect to breaches of their accuracy.

1233 Royal Charter on Self-Regulation of the Press, sched 3, cl 11.
1236 Wragg, A Free and Regulated Press (n 18) 276.
clause, this then is merely extending the scope of an existing remedy to cover the DNS clause.\textsuperscript{1238}

The final enforcement power is pecuniary; fines for breaching the codes. Both regulators currently have the power to fine regulated publications. However, they apply different thresholds for imposing fines. IPSO sets a higher threshold requiring serious and systemic violations, whereas IMPRESS can impose fines if a breach is serious \textit{or} systemic. This means that, unlike IMPRESS, IPSO cannot impose a fine for a single serious breach or multiple unserious ones.\textsuperscript{1239} Neither has used this power to date. They are also disincentivised from doing so due to the voluntary nature of membership. Any enforcement of the code that is perceived as heavy-handed could result in the publisher terminating their membership. One IMPRESS-regulated publisher did in fact threaten to do this in response to being the first—and only—publication found to breach the hate speech clause.\textsuperscript{1240}

To avoid this and to secure compliance under the voluntary regulatory model, the ultimate sanction for failing to pay a fine or otherwise comply with the code should be the termination of membership, a sanction that is already available to IPSO and IMPRESS. However, the difference here would be that these publications would then lose all the discretionary benefits that they acquired when they joined the regulatory body. These discretionary benefits, therefore, serve the dual purpose of incentivising entry into and compliance with the regulatory regime.

Finally, the question of what remedy ought to be deployed should be at the discretion of the regulator, and these should be proportionate to the nature and seriousness of the breach. For example, an article containing one DNS should be treated differently from one which is comprised entirely and overwhelmingly of them by an outlet with a poor record of compliance. Sanctions could be applied on a sliding scale ranging from a right to reply which keeps the offending parts of the publication up, to a correction that removes the offending parts and requires the outlet to publish the fact of the adjudication, to a fine where the breach is serious \textit{or} systemic. These fines should

\textsuperscript{1238} Council of Europe, Recommendation No. R (97) on ‘Hate Speech’, Principle 2; IMPRESS Standards Code, cl. 1.2; IPSO Editors Code, cl 1(iii).
\textsuperscript{1239} Wragg, A Free and Regulated Press (n 18) 262.
\textsuperscript{1240} ‘5Pillars Editors Respond to IMPRESS Adjudication on Discrimination Code’ (n 206).
also vary depending on whether the publication is a single or repeat offender. With respect to the voluntary regulatory model, expulsion from the regulator should only be a remedy of last resort where the publication refuses to comply with any of the other sanctions.

4. Compatibility of regulation with positive law

Having set out both the content and form of regulation, I will now consider whether these proposals would be compatible with the positive law. The co-regulatory model must be shown to be compliant as its mandatory nature means it amounts to a restriction of Article 10 by a public authority. It would therefore be amenable to judicial review under the Human Rights Act.²⁴¹

This examination is less relevant to voluntary regulation. I have set out why it does not count as a restriction on a convention right because membership is voluntary, and all of the proposals to incentivise membership set out above are benefits rather than burdens. Outlets that do not join a regulator are subject to the same legal treatment as any other publisher. Nevertheless, it is important to consider its compatibility in the event that it is amenable to judicial review. There are two reasons why this may be the case. The first is existing precedent which suggests that even regulators approved under the Royal Charter system are seen as performing a public function and may therefore be subject to judicial review.²⁴² Secondly, it may be argued that these incentives amount to restrictions in practice as they make unregulated publications less competitive than regulated ones and therefore risk the financial viability and survival of the former. If this is the case, it could be claimed that by allowing the creation of this system, the UK is in breach of its positive obligations to safeguard the press’s Article 10 rights.²⁴³

There is no neat answer to these questions in existing jurisprudence. Both models are applied in some form in states such as Austria, Belgium, Denmark, Germany, Ireland, and Poland. All impose duties and responsibilities not expected of the public, and all

²⁴¹ Human Rights Act 1998, s. 6; R (Gaunt) v Office of Communications [2011] EWCA Civ 692.
²⁴³ This positive rights view of human rights is now firmly entrenched in the case law of the ECtHR Gibbons (n 250) 20; Dimitris Xenos, The Positive Obligations of the State under the European Convention of Human Rights (1st edition, Routledge 2013).
have broader restrictions on speech, including restrictions on offensive content, that are more far-reaching than the targeted regulation of DNS proposed here.\textsuperscript{1244} All are signatories to the ECHR, but the ECtHR has not been called to determine whether these regimes are complaint with the convention. Equally—to the best of my knowledge—there are no domestic cases in the UK that examine this question. Absent this, definitive judgements cannot be made either way, and we must settle for predictions informed by the various tests deployed by the courts and related jurisprudence.

I will therefore scrutinise these proposals under the proportionality test utilised by the UK courts and the ECtHR. My analysis will primarily draw on the latter’s jurisprudence owing both to its rich body of cases relating to hate speech and because the domestic courts are both required to take their decisions into account and will rarely depart from them.\textsuperscript{1245} Finally, as the interference with Article 10 under both models is prescribed by law, I consider this requirement to be met and do not elaborate on it further, instead focusing on the other substantive stages of the test.

\textbf{a. Legitimate aim}

At this stage, the court looks to ascertain whether the restriction pursues an objectively justifiable policy. The UK courts have noted that the legitimate aim should be ‘sufficiently important to justify limiting a fundamental right’.\textsuperscript{1246} Restrictions on hate speech generally fall under the grounds of preventing disorder or crime, protecting morals and/or protecting the rights of others.\textsuperscript{1247} The court has recognised the various direct and indirect harms set out in this thesis as a legitimate basis for restrictions. In \textit{Aksu}, it noted that negative stereotypes amount to misrecognition which could lead to a loss of the victim groups, ‘feelings of self-worth and self-confidence of its members’.\textsuperscript{1248} In \textit{Erbakan},\textsuperscript{1249} \textit{Le Pen}\textsuperscript{1250} and \textit{Perinçek},\textsuperscript{1251} the courts noted that hate speech can create a climate of hatred and undermines the targeted group’s security

\begin{itemize}
\item \textsuperscript{1244} Heawood and Morris (n 37) 33–35.
\item \textsuperscript{1245} Human Rights Act 1998, s. 2 (1)(a); Roger Masterman, ‘Deconstructing the Mirror Principle’, \textit{The United Kingdom’s Statutory Bill of Rights} (British Academy 2013).
\item \textsuperscript{1246} R (Daly) v Secretary of State for the Home Department [2001] UKHL 26
\item \textsuperscript{1247} Féret v. Belgium [2009] Application No. 15615/07.
\item \textsuperscript{1248} Aksu v Turkey [2012] Application number 4149/04 and 41029/04.
\item \textsuperscript{1249} Erbakan v. Turkey [2006] Application number 59405/00.
\item \textsuperscript{1250} Le Pen v. France [2010] Application number 18788/09.
\item \textsuperscript{1251} Perinçek v. Switzerland [2013] Application no. 27510/08.
\end{itemize}
and dignity. The court has also specifically defined hate speech to include negative stereotyping.\textsuperscript{1252} Therefore, the restrictions on DNS in the press will likely be found to pursue a legitimate aim.

\textbf{b. Suitability}

Here the court determines whether there is a rational connection between the measure and the legitimate aim being pursued. The UK courts have stated the measure must be ‘carefully designed to achieve the objective’.\textsuperscript{1253} The ECtHR will likely find a policy that is unspecific and ineffective at achieving the objective to be disproportionate to the legitimate aim.\textsuperscript{1254} I have set out why—and demonstrated through applying it in practice—the DNS clause and guidance are both effective at catching the speech that causes harm. In terms of structure, both regulatory models clearly define which publications are included and excluded. The co-regulatory model’s mandatory nature and the voluntary model’s incentives are also effective at achieving compliance. The content and forms of regulation proposed are therefore suitable as they are rationally connected to the legitimate aim.

\textbf{c. Minimal impairment}

If a measure pursues a legitimate aim and is suitable, the courts will then consider whether it minimally impairs the right. The UK courts suggest that the means used ‘should be no more than necessary to accomplish the objective’.\textsuperscript{1255} The ECtHR asks whether there was a ‘pressing need’ for the interference.\textsuperscript{1256} A measure will fail to meet this test if there are less intrusive alternatives available to the state to pursue the same goals.\textsuperscript{1257} However, any alternative means must be as effective at achieving the

\begin{itemize}
\item[\textsuperscript{1252}] Gündüz v. Turkey [2003] Application number 35071/97, para 40.
\item[\textsuperscript{1253}] Huang v Secretary of State for the Home Department [2007] UKHL 11.
\item[\textsuperscript{1255}] Huang v Secretary of State for the Home Department [2007] UKHL 11.
\item[\textsuperscript{1256}] Handyside v. United Kingdom [1976] Application number 5493/7, para 48.
\item[\textsuperscript{1257}] Khosla (n 1254) 299.
\end{itemize}
objective otherwise comparisons become meaningless. The determination that it is as effective must be based on ‘concrete factual data’. Three questions flow from this. The first is whether there are alternative means of pursuing the objective which are equally effective and do not involve enacting content-based restrictions on speech. I have argued that there are good reasons to believe that such alternatives do not exist. A commonly proposed one is the more rigorous enforcement of an accuracy clause. However, I have set out why this has limited utility, as it is relevant only when DNS are presented as facts, rather than as opinions. Even in these limited cases relying on the clause is problematic. It may, for example, allow publications to present evidence proving the accuracy of certain negative stereotypes. This risks turning the proceedings into a forum to speculate about whether certain groups are predisposed to criminality or violence, statements whose ‘accuracy’ the committee is hardly capable of ascertaining.

On the other hand, we have evidence that the DNS clause is effective at catching this content. This is demonstrated both by using the clause to re-evaluate previous IMPRESS and IPSO decisions and from concrete factual data as a comparable provision in broadcasting has been effective at curbing discriminatory reporting in that medium.

The second question is whether there are less intrusive regulatory models which are equally effective. The voluntary model is certainly less intrusive than the mandatory co-regulatory model. I have also argued that robust incentives can be effective at securing membership. The court may therefore find that co-regulation is disproportionate because voluntary regulation is a less intrusive and equally effective means of pursuing the legitimate aim. However, this is still a speculative claim as we have no concrete factual data demonstrating that incentives are as effective at achieving universal coverage as mandatory membership. It is unclear how the court would resolve this. However, where there exist two or more reasonable alternatives

---

1258 The Oxford Handbook of Freedom of Speech (n 279) 184.
1260 Oster (n 24) 124; Barak (n 1259) 317.
1261 Longaker (n 1121) 132.
that are arguably equally as effective, the court would likely defer to the state to determine which among them they should adopt.\textsuperscript{1262}

The third question is whether there are enforcement powers that are both less intrusive and equally effective are ensuring compliance with the code. Arguably, the least restrictive of the proposed enforcement powers is the right of reply. While the court treats it as an interference with Article 10 as it infringes on editorial autonomy, it is a measure which is regarded as enhancing the public’s access to a plurality of opinions on matters of public interest.\textsuperscript{1263} Further, unlike a correction, the right of reply keeps the original article intact, and unlike a fine, the publication does not suffer a financial penalty beyond the opportunity cost of publishing different content in the space now reserved for the reply.

However, we have seen that to be an appropriate comparator, the less intrusive means must be as effective at limiting the harm caused by DNS as the other measure. There are reasons to doubt this is the case. I argue that a right of reply mediates the harmful effects of media stereotypes by providing for both intragroup contact and knowledge about the outgroup's beliefs. However, a more effective way of reducing harm is the prompt removal of such content to limit the prominence of DNS in public discourse or, better yet, if such content is not published at all in the first place. These are arguably better achieved through ordering corrections and fines. The former removes this content, and the latter targets the profit motivation which incentivises its production in the first place. I have noted that where reasonable alternatives exist, the court will likely defer to the state to determine which ones to adopt, and this also applies here. Further, the enforcement powers are not mutually exclusive but rather complementary. I have argued that the regulator should have all these powers available to them. Which ones they choose to deploy in a particular case should be determined both by the outlet’s compliance record and the severity of the breach. This is a question of proportionality in the narrow sense, which I now turn to.

\textsuperscript{1262} Fenwick and Phillipson (n 20) 99.
\textsuperscript{1263} See for example Melnychuk v. Ukraine [2005] Application no. 28743/03; Eker v. Turkey [2017] Application No. 24016/05.
d. Proportionality *strictu sensu*

By this point, the court has already determined that the state has a legitimate aim in restricting the right, is pursuing the legitimate aim through a suitable policy, and that there are no less intrusive means of resolving the conflict between the convention right and the legitimate aim. The court must then resolve the conflict by determining whether the state has struck a fair balance between press freedom and protecting minority groups from the harm of hate speech. This balancing test deployed in the final stage of the proportionality test is often referred to as proportionality *strictu sensu*.1265

In my examination of the justificatory theories of free speech and press freedom, I have relied on balancing to determine the proper scope of hate speech restrictions on the press. I have followed Alexy and Sumner who argue that human rights are optimisation requirements and that the goal of balancing is to locate the point at which ‘further gains in one of the values will be outweighed by greater losses in the other’.1266

The balancing test relied on by the UK courts and the ECtHR is framed in the same terms. In *National Union of Belgian Police*,1267 the court noted that it must determine whether ‘the disadvantage suffered by the applicant is excessive in relation to the legitimate aim pursued by the government’. Similarly, in *Sunday Times*, the court found that the state violated Article 10 because the interference did not ‘correspond to a social need sufficiently pressing to outweigh the public interest in freedom of expression’.1268 The UK courts have also framed the test in the same terms.1269

The similarities do not end here. While the previous stages of the proportionality test were factual, this stage is normative.1270 The weight (priority) given to either interest is informed by moral and political theory.1271 Activities which the court deems to be closely connected to the purposes of Article 10 are assigned considerable weight, and

1266 Sumner (n 332) 62; Alexy (n 336) 102.
1268 Sunday Times v UK (No.2) [1979] Application number 13166/87.
1270 Oster (n 24) 119; Fenwick and Phillipson (n 20) 101.
1271 Barendt, *Freedom of Speech* (n 6) 4.
restrictions on them are subject to greater scrutiny. This is similar to the heightened protection granted to speech at the core of a principle of free speech. The difference however is that the courts do not apply these arguments in the abstract, rather they are constrained by the text of the convention, specific legal rules, and precedent.\textsuperscript{1272}

I will now consider what theories inform the court’s approach to Article 10 and how they relate to the normative claims advanced in this thesis. Thereafter, I will examine the court’s precedent and legal doctrines, specifically the factors it considers in its balancing exercise and conclude whether these regulatory proposals are likely to be found to be compatible with the convention.

I have argued free speech theories suggest that the press ought to have more limited rights to engage in hate speech than the public. This view, supported by the SRT account of press freedom, is one of the justifications for the content and form of the broader regulation of press hate speech proposed above. Some suggest the ECtHR’s Article 10 jurisprudence is informed by a different normative framework, one which privileges media speech. Peter Coe notes that the presence of a media speaker ‘adds to the burden’ of justifying restrictions.\textsuperscript{1273} To support this view, commentators will point to the court’s decision in \textit{Jersild}, where the journalist’s conviction for aiding the disseminating of hate speech by an extremist group violated Article 10, yet the court found that members of the extremist group themselves could not benefit from the protection of the convention.\textsuperscript{1274} If this is the case, the imposition of broader hate speech restrictions on the press would \textit{prima facie} be disproportionate. But \textit{Jersild} must be contextualised. The court’s finding turned on a distinction between engaging in hate speech and dispassionately reporting another’s hate speech. It found that the purpose of the broadcast was the latter as it sought to ‘expose, analyse and explain’ the views of the extremist group to the public and raise social awareness on the use of racism.\textsuperscript{1275} The journalist would not have benefitted from any additional protection had he instead been propagating his own racist views.

\textsuperscript{1272} ibid; Frederick Schauer, ‘Must Speech Be Special?’ (1983) 78 Northwestern University Law Review 1284, 1305; \textit{The Oxford Handbook of Freedom of Speech} (n 279) 169–170.
\textsuperscript{1274} Jersild v Denmark [1994] Application number 15890/89.
\textsuperscript{1275} Ibid.
This fits in with the court’s general approach to Article 10. It has referenced various free speech theories, including the argument from individual self-fulfilment and the search for truth. However, the argument from democracy forms the bedrock of its Article 10 jurisprudence. The court assigns political speech, broadly understood as any speech contributing to debate on a matter of public concern, the greatest weight in the balancing test. It is therefore concerned with restrictions inhibiting the ‘dissemination of information relevant to the public’. The heightened protection that is sometimes extended to press speech is not because they have any unique rights as speakers but rather because they are institutions of mass speech capable of disseminating information at scale.

On the contrary, the ECtHR does accept the two central premises informing the claim that the press has more limited rights to engage in hate speech than the public. Firstly, it acknowledges that the press’s speech rights are protected solely for the benefit of the audience. In cases involving non-media speakers, the court refers to the individual’s right to impart ideas, but in cases involving media speakers the court frames it as a role, duty, or task to impart, as well as the ‘public right to know’. Secondly, it accepts that the media can be subject to more onerous duties and responsibilities than other speakers. This was most clearly laid out in Stoll, where the court found that:

*the safeguard afforded by Article 10 to journalists in relation to reporting on issues of general interest is subject to the proviso that they are acting in good faith and on an accurate factual basis and provide ‘reliable and precise’ information in accordance with the ethics of journalism.*

---

1276 Handyside v UK [1976] (Application number 5493/72); Fenwick and Phillipson (n 20) 39.
1278 Fenwick and Phillipson (n 20) 67.
Unlike other speakers, the media’s recourse to Article 10 is conditional on them being accurate, reliable and/or ethical.

Even if the court rejected these claims and instead adopted the libertarian view—that the press has the same rights to engage in hate speech as other speakers—it would still find these broader restrictions proportionate. The ECtHR is not an appellate court. It is ‘complimentary but subsidiary to national systems’.\textsuperscript{1282} In recognition of this, the court relies on the margin of appreciation doctrine to provide latitude and discretion to member states to implement the convention in light of their ‘own particular national circumstances and conditions’.\textsuperscript{1283} This means that even the greater scrutiny applied to restrictions on speech at the core of Article 10 still allows states significant discretion to determine how they formulate their hate speech laws. This margin of appreciation would be even wider with respect to voluntary regulation. This is because the court recognises that positive obligations are onerous and do not want to impose an ‘impossible or disproportionate’\textsuperscript{1284} burden, only requiring states to take ‘reasonable and appropriate measures’\textsuperscript{1285} to secure the convention rights. In effect, this means that the court will grant states a wide latitude to comply with positive obligations and will only find that a state has failed to do so when the interference is very serious.\textsuperscript{1286}

Finally, Article 10 allows for the restriction of a much broader category of hate speech than that which is caught under the stirring up offences.\textsuperscript{1287} While the stirring up offences are restricted to extreme language and incitement, the court in Gunduz found that states are justified in restricting ‘all forms of expression’ that ‘promote, incite or justify hatred based on intolerance’.\textsuperscript{1288} This is striking because it is the same definition of hate speech found in the Broadcasting Code.\textsuperscript{1289} The suggestion is that even if the
UK government chose to apply this broader definition to the rest of the populace, the court would not find such restrictions to be disproportionate.1290

The court also specifically refers to stereotyping as a form of regulable hate speech in Perinçek1291 and Aksu.1292 It even allows for restrictions on speech which injure religious feelings.1293 These are broader than those caught under the DNS clause, which is targeted at preventing harm, not offence. For example, in E.S, the court found that the applicant’s conviction for claiming that the Islamic prophet Muhammad was a paedophile was not disproportionate as it was likely to give rise to ‘justified indignation in Muslims’.1294 This statement would be protected under the DNS clause. The APPG guidance would only restrict such content if it was being used to characterise Muslims generally as ‘sex groomers’.1295

Three other factors relevant to proportionality are consensus, the penalties imposed and whether the restrictions inhibit the democratic functions of the media. The court will more intensely scrutinise restrictions on issues that are contra a ‘European consensus’ and conversely will apply a wide degree of latitude where such a consensus is absent.1296 We have seen that press regulation in Europe is diverse in both content and form, with varying models deployed in different countries. Therefore, the court is likely to grant the UK a wide degree of latitude to pursue either of the proposed models.

The second factor is the penalty imposed. The more serious the interference, the greater the intensity of review.1297 Custodial sentences in particular are carefully scrutinised.1298 So are prior restraints which seek to prevent publication rather than ‘punish’ it.1299 None of these factors are present in the proposed regulatory system nor in the enforcement measures. The court has in fact found similar enforcement

---

1290 Herz and Molnar (n 1040) 422.
1291 Perinek v Switzerland [2013] Application number 27510/08, para 97.
1292 Aksu v Turkey [2012] Application number 4149/04 and 41029/04.
1294 E.S. v Austria [2018] Application number 38450/12, para 57.
1295 Islamophobia Defined: The Inquiry into a Working Definition of Islamophobia’ (n 183) 56.
1296 Fenwick and Phillipson (n 20) 72–73.
1297 ibid 78–79.
1298 ibid (n 24) 138–139.
1299 ibid 123–130. However, they are not always strictly scrutinised Müller and others v. Switzerland [1988] (Application number 10737/84), para 36; Fenwick and Phillipson (n 20) 78–79.
measures to be proportionate. For example, it has found that bans on distributing publications containing hate speech as well as fines imposed on the authors are not per se disproportionate. In Eker, the court found that a right of reply that responded both to factual accuracies and statements of opinions was proportionate, noting that it ‘cannot be considered excessive or unreasonable’. 

The final factor is that restrictions that inhibit the press’s democratic functions are subject to intensive review. The proposed DNS clause caters for this by exempting the press from liability when it is acting as a public watchdog. However, the Jersild decision (discussed above) which immunises the press from liability when they are platforming other hate speakers is problematic as the protection is not conditional on them clearly distancing themselves from, or pushing back on, the speaker. This would appear to conflict with the platforming exemption to the DNS clause, where the exemption only applies if the publication challenges the hate speaker’s views. I have set out why this condition is necessary to avoid licensing the hate speakers with authority necessary for their speech acts to constitute harm. This point was alluded to in the dissenting judgments where some of the judges were critical of the journalist’s failure to push back, noting:

*the journalist responsible for the broadcast in question made no real attempt to challenge the points of view he was presenting, which was necessary if their impact was to be counterbalanced, at least for the viewers.*

More recent decisions suggest that the imposition of this requirement would be permissible under Article 10. In Sürek, the court emphasised the importance of the media avoiding becoming ‘a vehicle for the dissemination of hate speech and the promotion of violence’. We have also seen that in Stoll, the courts are more cognisant of the media’s influence and that their protection is subject to the proviso

---

1301 Altıntaş v Turkey [2020] Application no. 50495/08.
1302 Eker v. Turkey [2017] Application No. 24016/05), para 43.
1303 Von Hannover v Germany (No. 2) [2012] Applications no. 40660/08 and 60641/08, para 102; Axel Springer AG v Germany (No. 1) [2012] Application no. 39954/08, para 79.
1304 See Chapter 8
1305 Ibid Para 1 (Joint Dissenting Opinion Of Judges Gölcüklu, Russo And Valticos)
1306 Sürek and Özdemir v Turkey [1999] Applications number 23927/94 and 24277/94, para 63
that they are acting responsibly and ethically. The upshot of this is that the content and structure of restrictions imposed on the press are likely to pass muster under the Strasbourg court’s proportionality tests and would therefore be compatible with Article 10.

5. Conclusion

This chapter has proposed specific regulation that would be effective at addressing the issue of discriminatory reporting in the British press. I have shown that a DNS clause combined with group-specific guidance would effectively capture the problematic content while allowing regulators to engage in a contextualised analysis that is able to clearly distinguish between DNS and other controversial journalistic content. Relying on group-specific guidance also creates certainty for journalists who are able to reasonably forecast what content is caught by the clause. It allows us to prevent harm while granting leeway to the press to shock, provoke, exaggerate, and offend. Such guidance also serves a supplementary function as an educative tool informing journalists of the specific stereotypes that are used to encourage and justify prejudice against different groups.

In terms of structure, I have identified the criteria informing whom the regulation will apply to, including both the institutional press and forms of new media which resemble them in reach and communicative power. In earlier chapters, I have shown that the coercive application of broader regulation of hate speech in the press is normatively justified. In this chapter, I have shown enforcement of these standards through either co-regulation or voluntary regulation is compliant with the positive law. Therefore, the question of which to adopt is neither normative nor legal but rather informed by pragmatism.

Co-regulation benefits from mandatory membership and, therefore, universal coverage. It realises Leveson’s vision of a single regulator covering the entire industry. There are also institutional benefits of centralising regulation in terms of building expertise and a more consistent application of the code. However, there is unlikely to be political will for the imposition of a mandatory scheme. This is what Rawls would

1307 Stoll v Switzerland [2007] Application number 69698/01
term an ideal theory. It is the regulatory framework that we would apply in an ideal society unconstrained by the political infeasibility of such a measure.1308

Voluntary regulation coupled with a body that approves regulators which meet minimum standards (including the regulation of DNS) rather than a single mandatory regulator is likely to be more palatable to the press and, therefore, more politically expedient. We have seen that this only involves tweaks to, rather than an overhaul of the existing system. There are real concerns about a race to the bottom where the press will join a regulator that has the lightest touch to enforcing the DNS clause. However, this can be remedied by relying on the periodic review of approved regulators that is already part of the current charter system. Regulators that fail to enforce the clause should be de-listed, and their members would lose access to the discretionary benefits that come with that approval. Voluntary regulation is a non-ideal theory. It is what can realistically be accomplished given the political constraints that militate against the realisation of the ideal theory.

---

1308 Rawls distinguished between ideal and non-ideal theories in Rawls, A Theory of Justice (n 334); Rawls, Political Liberalism (n 876).
Conclusion

This thesis has set out to answer the following questions:

1. What kind of hate speech does the press publish and should it be restricted?

I have argued that hate speech by the British press remains an issue to date. While it does not appear in all—or even most—outlets, it is not an aberration relegated to fringe publications. It is a systemic issue that has attracted the attention of scholars, International and supranational human rights bodies, as well as in the most recent inquiry into press ethics by Lord Leveson.

I have pointed out that the genus of hate speech that is of concern is a form of negative stereotyping and stigmatisation which encourages and justifies prejudice through both essentialising and assigning an undesirable trait to racial, religious, and ethnic minorities; what I term dangerous negative stereotypes. In chapter 1, it was suggested that owing to legal proscriptions on overt hate speech and a liberalisation of social attitudes, hate speech including DNS, is now expressed through more moderate language. Overt DNS based on ideas of biological inferiority have been replaced by covert DNS and notions of cultural inferiority. I have argued that the latter is the kind of content that a regulator ought to have the power to deal with if Leveson's vision of restricting discriminatory reporting is to be realised.

I have also noted that current press regulators (IPSO and IMPRESS) do not proscribe this content. However, it is captured in the broadcast sector through a combination of a broader hate speech clause and Ofcom's willingness to engage in a contextual analysis that is sensitive to the group-specific nature of stereotypes. In chapter 3, I argued that there is no convincing rationale for this differential treatment of hate speech between the print and broadcasting press. This thesis has therefore sought to establish the normative justification for applying similar regulation to the print media and conceptualised how it would operate in practice.

In chapter 2, I argued that DNS should be restricted if they cause harm. Though harm is a necessary condition for regulation, it is not a sufficient one. In line with the balancing approach adopted in the UK (and in Europe generally), the justifiability of
restrictions on hate speech as well as their proper scope ought to be determined by balancing harm prevention with the need to protect fundamental rights, specifically press freedom and freedom of expression. Preventing all forms of racist propaganda will unduly impair free speech and press freedom while unfettered hate speech will cause unjustified harm to its targets. I have adopted a form of interest-based balancing appearing prominently in the work of Robert Alexy and applied by Leonard Sumner and Jan Oster to resolve line-drawing questions concerning hate speech by fringe extremists and by the media, respectively. The goal of this approach is to arrive at an optimal balance between these conflicting principles which is the ‘point at which further gains in one of the values will be outweighed by greater losses in the other’.  

I have relied on Frederick Schauer’s coverage/protection distinction to claim that speech at the core of a free speech principle, and activities in news gathering and publishing that fulfil the various democratic functions assigned to the press are both covered by and benefit from the greatest protection. Balancing in cases involving high-value speech/activities ought to proceed with a presumption in favour of protecting them, and any restrictions must be narrowly drawn. I have proceeded on the assumption that the stirring up offences which restrict only overt hate speech reflect an optimal balance in cases involving high-value speech/activities.

Under this approach, there are two circumstances where broader restrictions that regulate DNS would be compatible with either principle. The first is if the speech/activity in question is either low or no-value for purposes of either principle. Low-value speech/activities are covered by these principles but are at their periphery and benefit from minimal protection. I have referred to commercial speech under a free speech principle and celebrity gossip under a principle of press freedom as examples of low-value speech/activities. When balancing low-value speech/activities producing the same harm as high-value speech/activities, the priority should be to maximise the realisation of the competing principle of harm prevention and arriving at an optimal balance requires scaling up the constraints and imposing broader restrictions on the speech/activities in question. No-value speech/activities on the

---

1309 Sumner (n 332) 62; Alexy (n 336) 102.
other hand, are not even covered by these principles. As such, there is no balancing to be done, and the state has wide discretion to restrict them.

The second circumstance is where the speech/activity causes greater harm than that which justified restrictions on overt hate speech in the first place. I have noted that hate speech legislation is enacted to target fringe extremists, and when striking a balance between fundamental rights and harm prevention, legislators had in mind the lone hate speaker’s limited communicative power and capacity to cause harm. The optimal balance struck through this legislation is upset where owing to their greater communicative power, hate speech by the press causes greater harm than the speech of a lone hate speaker. I argue that balancing in such cases requires prioritising the satisfaction of the competing principle of harm prevention. In other words, all else being equal, optimal balancing requires scaling up the constraints on the speaker with greater communicative power (in this case, the press) to regulate their speech (DNS) causing this greater harm.

Put simply, I argue that these DNS should be restricted if either the press ought to have more limited rights to engage in hate speech than the public or if their hate speech produces greater harm than the overt hate speech of lone hate speakers.

2. What is the relationship between press freedom and free speech, and how can these principles be reconciled with restrictions on covert hate speech?

I have set out how the relationship between these two principles has been conceived in press freedom theory. I first considered the libertarian theory preferred by the industry which either sees the two as synonymous; that press freedom means the same thing as freedom of expression, or which recognises press freedom as a distinct principle but, for pragmatic reasons, argues they should be treated as if they are the same thing. The result is the same. Libertarian theory is predicated on an equivalence model where the press’s speech rights are identical to those possessed by the public.

I noted that there are two justifications in free speech theory for treating hate speech as high-value speech, the argument from self-disclosure and the argument from a citizen’s right to participate in democratic debate. Treating the press like other speakers means treating them as if they also have recourse to these justifications and
therefore treating their hate speech as high-value speech. As such, the only way that broader restrictions can justifiability be imposed on the press under the equivalence model is if the second circumstance above applies. This requires evidence that DNS are more harmful than the overt hate speech of lone hate speakers. I have examined various direct and indirect harms caused by hate speech arguing that we do have such evidence.

In terms of direct harms, I reviewed claims that overt hate speech causes psychological harm and severe emotional distress. I noted that the only forms of hate speech that cause these harms are those targeted at specific individuals, and often in interpersonal contexts. Such hate speech is already restricted in the general law by public order offences and in both press codes by provisions which restrict pejorative or prejudicial references to specific individuals. In most cases, diffuse hate speech (whether overt or covert or whether by lone hate speakers or by the press) does not cause psychological harm or severe emotional distress. These harms cannot, therefore, form a basis for regulating this genus of hate speech.

I reviewed claims that overt hate speech is a form of misrecognition which causes minorities to internalise the message of inferiority contained within such speech leading to a loss of self-esteem and even self-hatred. I have set out how this can cause the affected group to withdraw from or participate less in the formation of public opinion and/or self-censor by erasing their personhood to blend in with the majority. Both responses silence the victim and undermine their free speech interests. There is evidence that misrecognition also occurs from DNS published by the press, a phenomenon conceptualised as ‘stereotype threat’ in media effects literature. It is suggested that overt hate speech is more likely to silence than DNS. This is because a message of inferiority must be recognised by the victim in order to be internalised. Overt hate speech expresses this message explicitly and will always be recognised by minority groups. DNS by the press are less explicit. They are expressed through moderate language and may escape recognition by at least some members of the minority group.

However, misrecognition by the press is more problematic for two reasons. Firstly, they misrecognise at scale as they have greater reach and communicative power than most lone hate speakers. This means their DNS are more widely circulated and will
be seen by more members of the minority group. It is therefore a comparison between a higher likelihood of recognition by a much smaller number of the minority group and a (comparatively) lower likelihood of recognition by a much greater number of them. Secondly, cultural mistrust exacerbates the loss of self-esteem and occurs when minorities come to believe that the hate speaker’s views are representative of those held by the wider society. This leads them to be suspicious of and be more guarded in interactions with the majority, i.e. by self-censoring and retreating to ingroup interaction to a greater degree. I argue that because cultural mistrust springs from the minority groups’ perception as to how widespread the speaker's views are, it will be caused to a greater degree by the publication of DNS in a national or regional newspaper than by a reactionary diatribe by a fringe extremist.

A related question which this thesis did not consider is whether misrecognition amounts to a breach of the minority group’s right to private life and whether DNS can be restricted on this basis. This notion, developed in a recent line of ECtHR jurisprudence, was first set out in Aksu,\textsuperscript{1310} where the court found that private life includes a group’s sense of identity, self-worth, and self-confidence which can be undermined by stereotyping that reaches ‘a certain level’. The principle has been restated in Budinova and Chaprazov\textsuperscript{1311} and Behar and Gutman,\textsuperscript{1312} where the court referred to the problematic content as ‘extreme negative stereotyping’ i.e. DNS. This recent line of cases suggests that the court’s jurisprudence on negative stereotyping will increasingly be anchored on a state’s positive obligations under Article 8. As such, it is worth interrogating whether—and to what extent—a right to private life ought to encompass a group’s sense of self-identity, and how this ought to be balanced against freedom of expression and press freedom. This is a vast topic that could not have been reasonably explored or developed in this thesis, but one which is ripe for future research.

In chapter 5, I argued that the press’s greater capacity to harm is particularly clear when it comes to indirect harms. I examined the claims that overt hate speech causes discrimination and violence as well as undermines security and dignity (social

\textsuperscript{1310} Aksu v Turkey [2012] Application nos. 4149/04 and 41029/04.
\textsuperscript{1311} Budinova and Chaprazov v Bulgaria [2021] Application no. 12567/13 para 93-94
\textsuperscript{1312} Behar And Gutman v Bulgaria [2021] Application no. 29335/13, para 104-105
standing). It was noted that the stirring up offences are premised on a less imminent connection between speech and acts of violence and discrimination that treats these effects as downstream consequences, made possible by the creation of a climate of hatred of which hate speech is a part contributor. Further, the concern is not limited to objective security, as hate speech erodes subjective feelings of security by causing minorities to fear that they will be abused, shunned, or attacked when out in public. Finally, it lowers their dignity by undermining their positions as equals in rank and status and is a means by which social hierarchies are created and maintained. It was noted that all of these indirect harms are predicated on a single empirical claim; that hate speech can foment prejudice towards minority groups, and the evidence certainly suggests that the press is more effective at doing so than other speakers.

While media effects are indirect and mediated by several variables, there is evidence linking media stereotypes and prejudicial attitudes and/or behaviour towards minorities. Much of our knowledge of the content of stereotypes is from the media, as we rely on them to obtain information about cultures and groups which we do not have personal experience with. By drawing on Daniel Kahneman’s work on judgements and decision-making, media effects researchers have been able to demonstrate how this knowledge translates into prejudicial attitudes. It is argued that to conserve cognitive resources, people rely on media stereotypes as heuristics to make judgements about social groups. This leads them to develop attitudes and/or behaviour which conform to stereotypes, and this is made more likely when they are more vivid in their memories through repeated priming by the media. These effects are problematic because they are largely unconscious and automatic. Further, only a small subset of the population who are of low prejudice, have intragroup contact with the stereotyped group and possess subject matter knowledge (e.g. knowledge about Islam to reject distortions and stereotypes about the beliefs of Muslims) are inoculated from the attitudinal and behavioural effects of media stereotypes. DNS by the press are therefore particularly potent at fomenting prejudice, more so than lone hate speakers.

Finally, I have argued that unlike most lone hate speakers, DNS by the press does not just cause harm; they constitute it. By relying on J.L. Austin’s theory of speech acts, we have seen that hate speech also has illocutionary force and can constitute the act of authoritatively ranking minorities as inferior to others and therefore subordinating
them. Performing an authoritative speech act requires authority which lone hate
speakers rarely possess but which the press often does. The press possesses
epistemic authority, understood as credibility. Some outlets derive it from being
endorsed by those with practical authority, such as government officials writing for
them. All have epistemic authority licensed to them by their readers and amongst
those who find them credible. These speech acts enact harmful social norms,
providing authority for belief by authoritatively conveying the ‘fact’ that certain groups
are inferior, and authority for action directing us to treat them as such. These norms
provide strong incentives to comply and penalise non-compliance through social
censure. This makes them particularly pernicious as they are resistant to and
disincentivise responses such as counterspeech that are at least partly effective
against direct and indirect harms.

In conclusion, the empirically demonstrable direct and indirect causal harms and
constituted harms informed by speech act theory suggest that greater harm is caused
through DNS published by the press than is caused by the overt hate speech of lone
hate speakers. Therefore, the conditions under which broader regulation of DNS can
be justified under the equivalence model have been made out. Arriving at an optimal
balance requires scaling up the constraints to regulate the speech (DNS by the press),
causing and constituting these greater harms.

A potential objection to this is that there are other speakers including politicians,
bloggers, and social commentators with significant influence, following and credibility.
It is arguable that these figures equally have the communicative power necessary to
cause and constitute these various harms and therefore a limited focus on the press
would be erroneous. Many of these figures even have greater communicative power
than say a local news outlet.

Briefly, these factors are already recognised in the criteria for inclusion in the
regulatory framework. Influential bloggers, social commentators and other forms of
new media who run successful businesses producing news-related content would
qualify for membership. Further, most local outlets would not be included as most
would be microbusinesses that do not generate sufficient revenue to meet the criteria.
Secondly, under the positive law, the speaker’s position in society and influence are already used as criteria to determine whether speech constitutes incitement to hatred.\textsuperscript{1313} For example, politicians have a special responsibility to refrain from engaging in hate speech.\textsuperscript{1314}

Thirdly, we have seen that the press’s communicative power comes not only from their reach but because of their unique ability to synthesise, produce and reproduce content at scale. Further, these other figures do not often engage in their own original journalism but rather often provide commentary on the press’s original reporting. This makes the press unique in that what they decide to publish and how they frame their stories sets both the agenda and the terms of the political debate. As an example, it was noted that the inaccurate ‘Muslim foster care’ story reported in the Times\textsuperscript{1315} was widely circulated and generated significant public debate. Even the reporting of smaller outlets is often picked up, re-circulated and amplified by larger outlets and bloggers. An example is an inaccurate story suggesting that Rochdale Council has used taxpayer money to fund toilets to be used exclusively by Muslims.\textsuperscript{1316} The story first appeared in Anorak News,\textsuperscript{1317} an obscure online-only tabloid and was then picked up by the Daily Star, thereafter generating furore and controversy.\textsuperscript{1318}

I have also set out the second way the relationship between freedom and expression and press freedom has been conceived of in the literature, this being SRT which treats them as distinct principles, with the latter covering and protecting certain activities in news gathering and publishing that fulfil certain democratic functions assigned to the press. Here press claims for free speech and press freedom are scrutinised to determine whether they are justified by the rationales for protecting either.

\textsuperscript{1315} Sheppard v Daily Star (2010) PCC.
\textsuperscript{1318} Curran and Seaton (n 26) ch 10.
Under SRT, even if press hate speech does not cause greater harm, it can be subject to broader restriction if the first circumstance set out above pertains, i.e. if hate speech by the press is either low or no-value speech/activity under both principles.

I have examined the various rationale for treating hate speech as high-value speech and argued that the most persuasive accounts are the argument from self-disclosure and the argument from a citizen’s right to political participation. Both point to how hate speech restrictions erode political legitimacy. These restrictions must be narrowly drawn and targeted only at overt hate speech as the law must grant speakers room to disclose their values and participate in political debate using more moderate language, including through DNS. However, the press cannot rely on either account as both are predicated on the speech actually embodying the speaker’s values. Most of the press are commercial entities, and their output is best understood as profit-motivated rather than value-disclosing speech. Leveson and O’Neill were correct in noting that the press are not human beings with a need to self-express. Baker’s account for self-disclosure is either weaker (low-value) or does not apply to them entirely (no-value speech). Similarly, the press are not citizens entitled to engage in democratic deliberation, and this account simply does not apply to them. The press have no recourse to the only arguments justifying treating hate speech as high-value speech. They ought to have more limited rights than lone hate speakers to engage in hate speech, and broader restrictions imposed on them would be compatible with a free speech principle.

I have also argued that in most cases, hate speech is not high-value for purposes of a press freedom principle, as the various democratic functions assigned to the press do not require them to publish hate speech. However, there are limited circumstances where these functions do require them to do so, in which case such activities would be high-value. In chapter 8, I set out examples where performing the watchdog function may inadvertently lead to promoting DNS. Similarly, it is also high-value when the press acts as a platform for citizens to self-express or participate in democratic debate on their own terms. I have argued that regulation can be reconciled with press freedom in such cases by exempting the press from liability when they perform either of these functions. These exceptions are with respect to the watchdog role, conditional on the reporting being accurate and, in the case of the platforming role, conditional on
the press challenging a speaker’s views which is necessary in order to avoid lending the speaker legitimacy, credibility and epistemic authority required for their speech acts to constitute harm.

This thesis has demonstrated that broader regulation of hate speech in the press that goes beyond that imposed on the public is entirely compatible and can be reconciled with these prominent theories of press freedom.

3. What form should effective press-specific hate speech regulation take?

I have gone beyond examining the normative justification for such regulation and have (in chapter 9) considered how it could potentially be applied in practice, setting out both the content of the regulation and the structure of the regulatory body. In terms of content, I have argued that parliament ought to impose a duty on regulators requiring them to include a provision within their codes that effectively deals with DNS. I have suggested that to comply with this duty, regulators should consider including a specific clause covering DNS. They should also engage in a contextualised approach by relying on group-specific guidance to determine whether any stereotypes amount to DNS. Thereafter, I used the clause and relevant group-specific guidance to re-evaluate previous IPSO and IMPRESS decisions, demonstrating that they are both effective at catching the problematic content while not impinging on the press’s ability to publish content that offends, shocks, and disturbs.

The strength of this approach lies in its use of group-specific guidance, but this is also its greatest weakness. It can be difficult to arrive at a consensus or even compromise on what constitutes DNS against a group. Even the IHRA and APPG definitions which are widely adopted in the UK have generated controversy: The former for allegedly conflating criticism of the state of Israel and Antisemitism, the latter for conflating criticisms of religious doctrine with bigotry against Muslims. It is also problematic because these two definitions only cover a fraction of the various racial, ethnic, and religious groups represented in the UK, and an even smaller fraction of the protected characteristics. There is a need to agree upon and provide such guidance, and while this process might be contentious, it is not without solutions. In devising them, great weight should be placed on subject matter expertise. The question of what amounts
to Antisemitic, Islamophobic or homophobic stereotyping is a sociological one. There are institutions, scholars, and policymakers dedicated to examining how these phenomena manifest against specific groups historically and contemporarily in different societies. If they arrive at a consensus on a definition and guidance, their opinion ought to hold greater weight than those of non-subject matter experts, including the standpoint epistemology of the targeted groups themselves or the lay opinion of non-group members, including the regulated publications. Legal scholars and lawmakers should only contribute to assessing the guidance’s impact on fundamental rights and the technical details of its application.

4. To whom should it apply? And where would it fit in the current regulatory framework?

In terms of structure, I have focused on remedying the biggest weaknesses in current press regulation. This being the lack of universal coverage and independence from industry and either a lack of or unwillingness to exercise their powers to sanction. I have proposed a structure to ensure that regulation is effective, setting out whom it should apply to, the regulatory model, and the kind of enforcement powers regulators ought to have.

I argued that such regulation should apply to the press and have drawn on existing definitions of the press in statute and in proposed legislation which refers to entities that produce news-related material in the course of business, which is written by different authors, is subject to editorial control, and the entity has a turnover of more than £2,000,000 per annum. This covers both the forms of traditional media and new media that have the reach, influence and therefore the communicative power required to cause the harms that are of concern in this thesis. They are also the kinds of commercial outfits to which the speaker-based justifications for treating hate speech as high-value speech are less likely to apply.

I have considered two regulatory models. The first is co-regulation, where the entire industry is under the ambit of a single regulator. Membership is mandatory, and all publications that meet the criteria for inclusion are automatically covered. The second model is voluntary regulation which is merely a more robust form of the current Royal Charter system save that it is underpinned by statute for two reasons; the need for
statutory entrenchment to prevent undue government interference with the regulator and to create the incentives necessary to encourage membership. This model is a system for recognising regulators, and it is suggested that Leveson’s recommendation 38, which I have argued requires proscribing DNS, be turned into mandatory recognition criteria. As with the current system, regulators who fulfil the recognition criteria will be approved. Membership is voluntary and is encouraged through various economic incentives that only publications regulated by an approved regulator will benefit from. This includes a more favourable VAT status for their digital content and exemptions from the ambit of the upcoming Online Safety Bill and other similar legislation. It also includes measures aimed at strengthening the press’s negotiating power when bargaining with online platforms and news-aggregators for the use of their content, similar to the EU press publishers’ right and the Australia News Media Code. There are, of course, many other plausible incentive structures that could be employed and which should be explored. The focus should be on measures such as those set out above, which provide benefits to press outlets which join approved regulators but treat them no differently to other publishers if they choose not to. This is because incentives which subject the press to worse treatment than other publishers when they fail to join a regulator, what I have referred to as ‘burdens’, are legislative dead ends. Decades of wrangles over press regulation inform us that burdens proposed by Leveson, such as the doomed s. 40 of the Crimes and Court Act are unlikely to ever be enacted by a UK government.

In terms of enforcement, I have considered three possible powers. The first is the power to order corrections by requiring publications to remove the offending parts of the article. The second is the right to reply where the offending parts are maintained, but the publication is required to publish a reply by the affected group. This gives minorities access to a prominent platform which they would otherwise not possess. It also allows for intergroup contact and provides an avenue to educate the publication’s readership about a group’s cultures and customs, both of which I earlier noted are crucial in mediating the effects of media stereotypes and therefore minimising these indirect harms. The third is the power to issue fines. It is suggested that the approach by IMPRESS, which requires either serious or systemic breaches of the code, is more appropriate than that adopted by IPSO, which requires the breach to be serious and systemic. The extent of the fine should be left to the regulators to determine. However,
it should be proportionate to the nature of the breach, i.e. a single DNS in an outlet that has a good record of compliance should be treated differently from an article that is overwhelmingly composed of them in a publication with a poor record of compliance. The regulators should have the discretion to decide which of these enforcement powers they should deploy, and they could apply them on a scale based on the severity ranging from a right to reply, to a correction, to a correction and fine.

Compliance is easier to achieve in the co-regulatory model, as the outlets are mandated to be members and therefore have no choice but to comply with the regulator's decisions. However, voluntary regulation requires a third enforcement power, expulsion from the regulator. It is suggested that outlets which refuse to pay fines or comply with the regulator's directions should be expelled from the regulator and lose the various benefits that such membership affords. Similarly, the body in charge of recognising regulators should, as is the case with the current PRP, conduct periodic reviews to determine whether a regulator continues to meet the recognition criteria. Regulators who are unable to effectively proscribe DNS should be de-listed and become unapproved regulators, and this equally results in their members losing these benefits. These benefits serve the dual purpose of incentivising entry into and compliance with the regulatory regime.

Finally, I have demonstrated how both models and these enforcement powers are compliant with Article 10. Both serve the legitimate aim of harm prevention. They are suitable in that they are effective and are carefully designed to achieve this objective. They minimally impair the right by creating certainty, allowing the press to publish offensive content, and avoiding creating a chilling effect. The exemptions for the watchdog and platforming functions mean that these restrictions do not prevent the press from fulfilling their democratic functions. Both would equally be proportionate in the narrow sense because the ECtHR has granted states a wide margin of appreciation in cases involving hate speech and has found restrictions on hate speech that go further than those proposed here to be compliant with Article 10.

However, I have ultimately settled on voluntary regulation as the preferable model because it is more politically feasible, as the press has and will resist—and the state
will continue to entertain such resistance to\textsuperscript{1319}—any mandatory model. Voluntary regulation is an incremental change, as it only involves a tweak to the existing structure. This thesis argues that there is no normative or legal barrier preventing reform, and the voluntary model eliminates the political barrier as well. We have the tools necessary to protect groups from harm while preserving a free press. It is incumbent on us to use them.

Even if this more incremental improvement is rejected and the current system is maintained, the proposals set out in this thesis can still be used to develop training material and guidance for editors and journalists. The purpose of such material would be to educate and sensitise them to the various harms caused by DNS and relay to them the importance of avoiding publishing such content. There is precedent for this as IPSO already guides publications on how to report on matters concerning transgender individuals\textsuperscript{1320} as well as Muslims\textsuperscript{1321} and even points out the importance of avoiding stereotyping these groups. However, these guidances are generic in that they do not set out group-specific examples of harmful stereotypes. The regulators should be encouraged to update their existing guidances to set out the specific stereotypes which are likely to foment prejudice against different groups, and to provide similar guidance for the other protected characteristics.

\begin{footnotesize}
\textsuperscript{1319} Wragg, \textit{A Free and Regulated Press} (n 18) 67.
\textsuperscript{1320} ‘Guidance on Researching and Reporting Stories Involving Transgender Individuals’ (IPSO 2016) <ipso.co.uk/media/1275/guidance_transgender-reporting.pdf> accessed 14 January 2020.
\end{footnotesize}
Bibliography

1. Books

3. ———, Reconfiguring Islamophobia (Palgrave Macmillan 2020)
5. Amos M, Harrison J and Woods L (eds), Freedom of Expression and the Media (Brill Nijhoff 2012)
7. Appiah KA, ‘What’s Wrong with Defamation of Religion?’ in Michael Herz and Peter Molnar (eds), The Content and Context of Hate Speech: Rethinking Regulation and Responses (Cambridge University Press 2012)
8. Austin JL, How to Do Things with Words (Harvard University Press 1975)
12. ———, Media, Markets, and Democracy (Cambridge University Press 2009)
31. ———, *Media and Democracy* (1st edn, Routledge 2011)


41. ——, ‘Foreword’, *Extreme Speech and Democracy* (Oxford University Press 2009)

42. ——, ‘Reply to Jeremy Waldron’ in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press 2012)

43. ——, *Taking Rights Seriously* (Bloomsbury Academic 2013)


47. ——, *The Moral Limits of the Criminal Law: Volume 2: Offense to Others* (Oxford University Press 1988)


57. ———, *Speech Matters: Getting Free Speech Right* (University of Queensland Press 2011)


64. Habib S, *Islam and Homosexuality* (Praeger 2009)


66. Hare I (ed), *Extreme Speech and Democracy* (Oxford University Press, USA 2011)


72. ———, *Hate Speech and Democratic Citizenship* (Oxford University Press 2016)


76. Heyman SJ, *Free Speech and Human Dignity* (Yale University Press 2008)

77. ———, ‘Hate Speech, Public Discourse, and the First Amendment’ in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy* (Oxford University Press 2009)


85. Karakoulaki M, *Critical Perspectives on Migration in the Twenty-First Century* (Laura Southgate and Jakob Steiner eds, E-International Relations 2018)


100. Malik M, ‘Extreme Speech and Liberalism’, *Extreme Speech and Democracy* (Oxford University Press 2009)


112. Mutz DC, Brody RA and Sniderman PM (eds), *Political Persuasion and Attitude Change* (University of Michigan Press 1996)


120. ———, ‘Is There a Case for Banning Hate Speech?’ in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press 2012)


129. ———, Democracy, Expertise, and Academic Freedom: A First Amendment Jurisprudence for the Modern State (Yale University Press 2012)

130. Rae Langton, Haslanger S and Anderson L, ‘Language and Race’ in Gillian Russell and Delia Graff Fara (eds), The Routledge companion to philosophy of language (Routledge 2013)


133. ———, Justice as Fairness: A Restatement (2nd edn, Harvard University Press 2001)


138. ———, Media Law (Hart Publishing 2018)


140. Salmon WC, Causality and Explanation (Oxford University Press 1998)


149. Stefancic J and Delgado R, Must We Defend Nazis?: Why the First Amendment Should Not Protect Hate Speech and White Supremacy (NYU Press 2018)
150. Sumner LW, The Hateful and the Obscene: Studies in the Limits of Free Expression (University of Toronto Press 2016)
153. Tambini D, Media Freedom (Polity 2021)
156. The Oxford Handbook of Freedom of Speech (Oxford University Press 2021)
157. Tittley G, Racism and Media (SAGE 2019)
159. Waldron J, ‘Hate Speech and Political Legitimacy’ in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press 2012)

160. ———, *The Harm in Hate Speech* (Harvard University Press 2014)


2. Articles

1. Allington D and Joshi T, ‘Antisemitism and the “Alternative Media”’ (Kings College London 2021)

2. Altman A, ‘Liberalism and Campus Hate Speech: A Philosophical Examination’ (1993) 103 Ethics 302


5. Andreotti O and others, Journalism at Risk: Threats, Challenges and Perspectives (Council of Europe 2015)


26. Blázquez FJC and others, ‘On-Demand Services and the Material Scope of the AVMSD’ (European Audiovisual Observatory 2016) IRIS Plus 2016-1


44. Chivers T, ‘Who Owns the UK Media?’ (Media Reform Coalition 2021)


49. Coe P, ‘Redefining “Media” Using a “Media-as-a-Constitutional-Component” Concept: An Evaluation of the Need for the European Court of Human Rights to Alter Its Understanding of “Media” within a New Media Landscape’ (2017) 37 Legal Studies 25


64. Fasoli F and others, ‘Not “Just Words”: Exposure to Homophobic Epithets Leads to Dehumanizing and Physical Distancing from Gay Men’ (2016) 46 European journal of social psychology 237


71. ——, ‘Evidencing the Harms of Hate Speech’ (2016) 22 Social Identities 324


81. Hare I, ‘Crosses, Crescents and Sacred Cows: Criminalising Incitement to Religious Hatred’ [2006] Public Law 521


86.——, ‘Hate Speech and the Normative Foundations of Regulation’ (2013) 9 International Journal of Law in Context 590


89. Howard JW, ‘Free Speech and Hate Speech’ (2019) 22 Annual Review of Political Science 93


118. Lewis D, ‘Scorekeeping in a Language Game’ (1979) 8 Journal of Philosophical Logic 339


123. McGonagle T, ‘Minorities and Online “Hate Speech”: A Parsing of Selected Complexities’ (2012) 9 European Yearbook of Minority Issues Online 419


133. Müller K and Schwarz C, ‘Fanning the Flames of Hate: Social Media and Hate Crime’ (2021) 19 Journal of the European Economic Association 2131


142. Phoenix A, ‘Somali Young Women and Hierarchies of Belonging’ (2011) 19 YOUNG 313


144. ———, ‘Participatory Democracy and Free Speech’ (2011) 97 Virginia Law Review Online 477


148. ——, ‘Intergroup Contact, Media Exposure, and Racial Attitudes’ (2013) 42 Journal of Intercultural Communication Research 54


155. Saleem M and others, ‘Exposure to Muslims in Media and Support for Public Policies Harming Muslims’ (2017) 44 Communication Research 841


159. ——, ‘Comment on Baker’s Autonomy and Free Speech’ (2011) 27 Constitutional Commentary 319


162. ——, ‘Can Rights Be Abused?’ (1981) 31 The Philosophical Quarterly 225
163. ______, ‘Must Speech Be Special?’ (1983) 78 Northwestern University Law Review 1284
164. ______, ‘The Phenomenology of Speech and Harm’ (1993) 103 Ethics 635
171. Simpson RM, ‘Dignity, Harm and Hate Speech’ (2013) 32 Law and Philosophy 701
175. Sorial S, ‘Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement’ (2015) 34 Law and Philosophy 299


180. Steinhoff U, ‘Really Just Words: Against McGowan’s Arguments for Further Speech Regulation’ [2022] Philosophia


182. Sumner LW, ‘Reviewed Work(s): Speech and Harm: Controversies over Free Speech by Ishani Maitra and Mary Kate McGowan’ (2013) 39 Social Theory and Practice 710


205. Wright RG, ‘Racist Speech and the First Amendment’ (1988) 9 Mississippi College Law Review 1


3. Table of cases

**European Court of Human Rights/European Commission of Human Rights**

2. Altıntaş v Turkey [2020] Application no. 50495/08
3. Axel Springer AG v Germany (No. 1) [2012] Application no. 39954/08
10. E.S. v Austria [2018] Application number 38450/12
11. Eker v Turkey [2017] Application No. 24016/05
12. Erbakan v Turkey [2006] Application number 59405/00
15. Gündüz v Turkey [2003] Application number 35071/97
21. Müller and others v Switzerland [1988] (Application number 10737/84)
22. Murphy v Ireland [2003] Application no. 44179/98 77
27. Perinçek v Switzerland [2013] Application number 27510/08

286
31. Stoll v Switzerland [2007] Application number 69698/01
32. Sunday Times v UK (No.2) [1979] (Application number 13166/87)
34. Thoma v Luxembourg [2001] Application number 38432/97
36. Times Newspapers Ltd v. United Kingdom [2009] Applications nos. 3002/03 and 23676/03
37. Von Hannover v Germany (No. 2) [2012] Applications no. 40660/08 and 60641/08
38. X and Church of Scientology [1979] Application number 7805/77

United Kingdom

3. Gulati & Ors v MGN Ltd [2015] EWHC 1482 (Ch)
7. Prolife Alliance v BBC [2003] UKHL 23
8. R (Daly) v Secretary of State for the Home Department [2001] UKHL 26
9. R (Gaunt) v Office of Communications [2011] EWCA Civ 692
10. R v Bitton [2019] EWCA Crim 1372
11. R v Bonehill-Paine, Sentencing Remarks of Mr Justice Spencer (2016)
United States

4. Schenck v United States, 249 US 47, 52 (1919)

Canada


Decisions by media regulators

1. **PCC**
   c. Sharon v The Independent [2003] PCC

2. **IPSO**
   a. Elgy v The Sun [2016] IPSO 17562-17
   b. Greer v The Sun [2015] IPSO 02741-15
   c. Manji v the Sun [2016] IPSO 05935-16
   d. Tower Hamlets Borough Council v The Times [2018] IPSO 20480-17
   e. Trans Media Watch v The Sun [2015] IPSO 00572-15

3. **IMPRESS**
   a. 5 Pillars [2020] 303/2020 IMPRESS 15
   b. The Canary and The Skwawkbox [2021] IMPRESS 364/2021

4. **OFCOM**
   c. The Sun Video case (2011)