Regulating Extreme Speech in the Digital Age: A Comparative Analysis of European and American Approaches

BECHTOLD, ELISABETH,VIRGINIA

Regulating Extreme Speech in the Digital Age:

A Comparative Analysis of European and American Approaches

Eliza Bechtold

Abstract

While the notion that freedom of speech is a well-established and valuable right is uncontested in liberal democracies, significant disagreements exist with respect to how it is conceptualised in legal frameworks and the extent to which government restrictions on the message or content of expression are considered legitimate. These differences raise significant challenges that are the subject of longstanding debates. In the digital age, advances in technology have transformed these debates by, among other things, profoundly altering the ways in which people communicate with one another and how governments communicate with the public. These advances have opened new pathways for communicating in public discourse while presenting new challenges and opportunities for speech regulation.

This thesis critically examines the most challenging and significant contemporary free speech questions raised by three types of extreme speech - 'hate speech', terrorist-related expression, and disinformation from state actors - through a comparative constitutional inquiry of the regulatory approaches of Europe and the United States. While the problems confronting the Europe and the United States relating to the causes and consequences of online extreme speech are similar, the approaches vary in significant and meaningful ways. Thus, situating such an inquiry within a broader comparative analysis provides for richer and more nuanced observations than result from inquiries focusing on a single jurisdiction. Additionally, engaging in an analysis of these types of extreme speech in a single volume highlights the unique harms and regulatory challenges flowing from each type of speech while illuminating interrelated issues from which broader themes, lessons, and connections emerge. In so doing, this thesis offers an original contribution to the broader discourse concerning the appropriate limits on freedom of expression in liberal democracies in the digital age.

Regulating Extreme Speech in the Digital Age

A Comparative Analysis of European and American Approaches

Eliza Bechtold

Thesis submitted for the Degree of

Doctor of Philosophy

Durham Law School

Durham University

September 2021

Table of Contents

Declaration of Copyright

"The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged".

Certain portions of Chapter 4 have already been published in a book chapter: Eliza Bechtold and Gavin Phillipson, 'Glorification of Censorship? Anti-Terror Law, Speech, and Online Regulation' in Adrienne Stone and Frederick Schauer (eds) *The Oxford Handbook on Freedom of Speech* (OUP 2021). Part of this chapter was written by Gavin Phillipson. No part of the work of Gavin Phillipson is reproduced in this thesis. All of the work in this thesis is that of the author.

Acknowledgements

Over the past three years, I have benefitted from the support and guidance of many people. First and foremost, Professor Gavin Phillipson, who approached his supervisory responsibilities with gusto and continued to support me well after he was no longer contractually obliged to do so. Additionally, Dr Eleni Frantziou went above and beyond her role as secondary supervisor by providing invaluable feedback and enthusiastic support throughout the research and writing process. I am grateful to both Professor Phillipson and Dr Frantziou for their support and guidance. I am also thankful for the guidance and helpful feedback provided by my primary supervisor, Professor Helen Fenwick.

I have further benefitted from the support of the fantastic community of early career researchers at Durham Law School. These include Dr Kyle Murray, Dr Dimitrios Kagiaros, Dr Benedict Douglas, and Dr Elizabeth O'Loughlin. Each of these individuals took time out of busy teaching and research schedules to answer (many) questions, review material, and/or provide guidance regarding the writing process. The efforts of these individuals made an indelible impact on the quality of this thesis and I am extremely appreciative.

Dedication

For my parents, Janet and George, for instilling in me the belief that any goal is within reach, encouraging me to follow my own path (no matter how circuitous), and for cheering me on every step of the way.

Table of Authorities

Cases

European Court of Justice

Strasbourg (European Court of Human Rights)

United States of America

United States v Schwimmer, 279 US 644 (1929)

United States v Sherwood, 312 US 584 (1941)

United States v Stevens, 130 SCt 1577 (2010)

Virginia v Black, 538 US 343 (2003)

Walker v Texas Div, Sons of Confederate Veterans, Inc, 576 US 200 (2015)

Ward v Rock Against Racism, 491 US 781 (1989)

Warner Cable Communications, Inc v City of Niceville, 911 F2d 634 (11th Cir 1990)

Watts v United States, 394 US 705 (1969)

West v Atkins, 487 US 42 (1988)

Whitney v California, 274 US 357 (1927)

Williams v United States, 71 F3d 502 (5th Cir 1995)

Wilson v Libby, 535 F3d 697 (DC Cir 2008)

Wood v Georgia, 370 US 375 (1962)

Zeran v Am Online, Inc, 129 F3d 327 (4th Cir 1997)

Legislation

Council of Europe

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR)

Council of Europe Convention on the Prevention of Terrorism (Warsaw) CETS No.196

England and Wales

Human Rights Act 1998

Terrorism Act 2006

Counter-Terrorism and Border Security Act 2019

European Union

Charter of Fundamental Rights of the European Union [2012] OJ C326/391

Treaty establishing a Constitution for Europe [2004] OJ C310/261

Decisions

Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law [2008] OJ L328

Regulations

Council and Parliament Regulation (EU) 2021/784 of 29 April 2021 on addressing the dissemination of terrorist content online [2021] OJ L172/79

France

LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information.

Germany

Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG) (Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act))

Grundgesetz (Basic Law for the Federal Republic of Germany)

Strafgesetzbuc (German Criminal Code)

Iceland

General Penal Code No. 19/1940 (*Almenn hegningarlög)*

United States of America

Anti-Terrorism Act 18 USC

Biased Algorithm Deterrence Act of 2019 (HR 492)

Communications Act of 1934 (47 USCA)

Communications Decency Act of 1996 (CDA) 47 USC

Federal Tort Claims Act of 1946 28 USC

International Treaties and Documents

International Covenant on Civil and Political Rights (adopted 16 December 1996, entered into force 23 March 1976) 999 UNTS 171

International Convention on the Elimination of All Forms of Racial Discrimination (adopted and opened for signature by General Assembly resolution 2106 (XX) of 21 December 1965, entered into force 4 January 1969)

Official Documents

Council of Europe

Akandji-Kombe J-F, 'Positive Obligations under the European Convention on Human Rights: A Guide to the implementation of the European Convention on Human Rights (Human Rights Handbooks No 7)' (Council of Europe, 2007)

Cilevičs B, 'Should politicians be prosecuted for statements made in the exercise of their mandate? (Provisional Version)' (Council of Europe Committee on Legal Affairs and Human Rights, 7 June 2021)

Council of Europe, 'Supervision of the Execution of Judgments and Decisions of the European Court of Human Rights 2017: 11th Annual Report of the Committee of Ministers' (March 2018)

— — 'Human Rights Comment' (12 April 2018) <https://www.coe.int/en/web/commissioner/-/misuse-of-anti-terror-legislation-threatens-freedom-of-expression> accessed 1 April 2019

— — 'European Union accessed to the European Convention on Human Rights - Questions and Answers' <https://www.coe.int/en/web/portal/eu-accession-echr-questions-and-answers> accessed 2 May 2021

— — 'News of the European Commission against Racism and Intolerance' (27 February 2020) <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/-/ultra-nationalism-antisemitism-anti-muslim-hatred-council-of-europe-s-anti-racism-commission-raises-alarm-over-the-situation-in-europe> accessed 5 November 2020

— — 'Interpretive Mechanisms of ECHR case-law: the concept of European Consensus' <https://www.coe.int/en/web/help/article-echr-case-law> accessed 9 July 2021

ECRI Institute, 'General Policy Recommendation No 15 on Combating Hate Speech' (ECRI, 21 March 2016) <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01> accessed 19 September 2020

European Court of Human Rights, ‘Guide on Article 17 of the European Convention on Human Rights: Protection of abuse of rights’ (European Court of Human Rights, updated on 31 August 2019) available at <https://www.echr.coe.int/Documents/Guide_Art_17_ENG.pdf> accessed 10 July 2020.

Venice Commission, ‘Russian Federation Opinion on the Draft Amendments to the Constitution (As Signed by the President of the Russian Federation on 14 March 2020) Related to the Execution in the Russian Federal of Decisions by the European Court of Human Rights (18 June 2020) Opinion No 981/2020 (Venice Commission, 2020)

— — ‘Report on the Relationship between Freedom of Expression and Freedom of Religion: the Issue of Regulation and Prosecution of Blasphemy, Religious Insult and Incitement to Religious Hatred’ (Venice Commission, 23 October 2008) <https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2008)026-e> accessed 10 July 2020

— — ‘Report of the Implementation of International Human Rights Treaties in Domestic Law and the Role of the Courts’ (Venice Commission, 8 December 2014) CDL-AD(2014)036

— — ‘Report on the Relationship between Freedom of Expression and Freedom of Religion: the Issue of Regulation and Prosecution of Blasphemy, Religious Insult and Incitement to Religious Hatred (Venice Commission, October 2008)

Registrar of the ECtHR, ‘Breach of the right to freedom of expression in the context of an electoral campaign (Press Release)’ (Registrar of the ECtHR, 2019) ECHR 284

Steering Committee on Media and Information Society (CDMSI), ‘Recommendation CM/rec(2014)6 of the Committee of Ministers to member States on a guide to human rights for Internet users – Explanatory Memorandum’ CM(2014)31 (16 April 2014)

Wardle C and Derakshan H, ‘INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making’ (Council of Europe Report DGI(2017)19’ (Council of Europe, 27 September 2017)) <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c> accessed 30 June 2021

European Union

Alaphilippe A and others, ‘Automated tackling of disinformation’ (Publications Office of the European Union, 2019) <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU(2019)624278_EN.pdf> accessed 21 May 2021

Bayer J and others, ‘Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States (Directorate-General for External Policies)’ (Publications Office of the European Union, 2019)

Directorate-General for Internal Policies of the Union, ‘The European legal framework on hate speech, blasphemy and its interaction with freedom of expression’ (2015)

<https://www.europarl.europa.eu/RegData/etudes/STUD/2015/536460/IPOL_STU(2015)536460_EN.pdf > accessed 18 December 2020

Directorate-General of Human Rights and Rule of Law, Recommendations and Declarations of the Committee of Ministers of the Council of Europe in the field of media and information society' (Directorate General of Human Rights and Rule of Law, 2016) 76 - 78 outlining Recommendation No R (97) 20 of the Committee of Ministers to member states on "hate speech" (adopted 30 October 1997)

Drachenberg R and Phillips E, 'The Twitter activity of members of the European Council: A content analysis of EU leaders' use of Twitter in 2019-20' (Publications Office of the European Union, 8 January 2021) <https://www.europarl.europa.eu/RegData/etudes/STUD/2021/654200/EPRS_STU(2021)654200_EN.pdf> accessed 20 May 2021

European Commission, Europeans' attitudes toward security' (*European Commission*, 2017) Special Eurobarometer 464b <https://europa.eu/eurobarometer/surveys/detail/1569> accessed 30 September 2019

— — 'State of the Union 2018: Commission proposes new rules to get terrorist content off the web' (*European Commission Press Release,* 2 September 2018) <http://europa.eu/rapid/press-release_IP-18-5561_en.htm> accessed 3 April 2019.

— — 'EU Internet Forum: Civil Society Empowerment Programme' <https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network/civil-society-empowerment-programme> accessed 8 February 2020

— — 'Tackling online disinformation' <https://digital-strategy.ec.europa.eu/en/policies/online-disinformation> accessed 1 July 2021.

— — 'The Digital Services Act package' <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> accessed 15 July 2021

— — 'Fighting Terrorism Online: Public-private sector cooperation as important as ever at the fourth EU Internet Forum (European Commission Statement, 5 December 2018) <https://ec.europa.eu/commission/presscorner/detail/en/statement_18_6681> accessed 10 February 2019

— — 'The EU's accession to the European Convention on Human Rights: Joint statement on behalf of the Council of Europe and the European Commission (Commission Press Corner, 29 September 2020) <https://ec.europa.eu/commission/presscorner/detail/en/statement_20_1748> accessed 13 March 2021

— — 'A multi-dimensional approach to disinformation: Report of the independent High-level Group on fake news and online disinformation' (Publications Office of the European Union, 2018) <http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271> accessed 1 July 2021

— — 'The EU Code of conduct on countering illegal hate speech online: The robust response provided by the European Union' (ST 12522/19 COR 1, 2 October 2019) <https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play__0.pdf> accessed 20 June 2021.

— — ‘Code of Conduct – Illegal online state speech questions and answers’ (EC Fact Sheet, June 2016) <https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf> accessed 19 September 2020

— — ‘Answer given by Vice-President Mogherini on Behalf of the Commission (Question Reference E-005122/2016)’ (Parliamentary Questions, 15 September 2016)

European Parliament, ‘Legislative Train 06.2021, Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC / After 2020-9’ (*European Parliament*, 2021) <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act> accessed 10 July 2021

European Union Agency for Fundamental Rights, ‘Proposal for a Regulation on preventing the dissemination of terrorist content online and its fundamental rights implications: Opinion of the European Union Agency for Fundamental Rights’ (*Publications Office of the European Union*, 2019)

Fletcher R and Jenkins J, ‘Polarisation and the news media in Europe’ (Publications Office of the European Union, 2019) <https://op.europa.eu/en/publication-detail/-/publication/914380a0-8e62-11e9-9369-01aa75ed71a1/language-en> accessed 18 May 2021

Hardt S, ‘Parliamentary Immunity in a European context (Directorate General for Internal Policies (Policy Department C: Citizens’ Rights and Constitutional Affairs))’ (Publications Office of the European Union, 2015) <https://publications.parliament.uk/pa/jt199899/jtselect/jtpriv/43/43ap21.htm> accessed 15 May 2021

Lewandowsky S and others, ‘Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making (EU Joint Research Centre)’ (Publications Office of the European Union, 2020) doi:10.2760/709177

Marzocchi O, ‘The protection of Article 2 TFEU values in the EU’, European Parliament Fact Sheet (April 2021) <https://www.europarl.europa.eu/factsheets/en/sheet/146/the-protection-of-article-2-teu-values-in-the-eu> accessed 2 July 2021

Panizza R and Pavy E, ‘Handbook on the incompatibilities and immunity of the Members of the European Parliament’ (Publications Office of the European Union, March 2020)

PRISM Project, ‘Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU Wide Study and National Assessments’, Prism Project <https://ec.europa.eu/migrant-integration/librarydoc/hate-crime-and-hate-speech-in-europe-comprehensive-analysis-of-international-law-principles-eu-wide-study-and-national-assessments> accessed 15 September 2020

Raffaelli R and Sy SS, ‘The Immunity of Members of the European Parliament (Directorate-General for Internal Policies, Policy Department C: Citizens’ Rights and Constitutional Affairs)’ (Publications Office of the European Union, 2014)

United Kingdom

‘Concluding observations on the twenty-first to twenty-third periodic reports of United Kingdom’ (CERD/C/GBR/CO/21-23, 26 August 2016)

‘Online Harms White Paper’ (*HM Government*, April 2019) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf> accessed 2 May 2019

United Nations

Office of the United Nations High Commissioner for Human Rights, ‘Joint Open letter on concerns about the global increase in hate speech’ (23 September 2019) <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25036&LangID=E#:~:text=OHCHR%20%7C%20Joint%20open%20letter%20on,global%20increase%20in%20hate%20speech&text=We%20are%20alarmed%20by%20the,their%20rights%2C%20in%20numerous%20countries.> accessed 2 November 2020

United Nations, ‘United Nations Strategy and Plan of Action on Hate Speech’ (UN, May 2019) 2 <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf > accessed 25 June 2021

— — ‘UN Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences’ (September 2020) <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf> accessed 4 December 2020

United Nations General Assembly, ‘Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (7 September 2012) UN Doc A/67/357

— — ‘Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (UNGA, 9 October 2019) UN Doc A/74/486

— — ‘Resolution on Israeli settlements in the Occupied Palestinian Territory, including East Jerusalem, and the occupied Syrian Golan’ (UNGA, 23 December 2016) UN Doc A/RES/71/97

— — ‘Human Rights Council Report on the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms high countering terrorism’ (*UNGA,* 1 March 2019) UN Doc A/HRC/40/52

— — ‘Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms’ (UNGA, 9 October 2019) UN Doc A/74/486

United Nations High Commissioner for Refugees, Addendum to the Annual report of the United Nations High Commissioner for Human Rights’ (UNHCR, 11 January 2013) UN Doc A/HRC/22/17/Add.4.

— — ‘One-pager on “incitement to hatred”’ (UNHCR) <https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_threshold_test.pdf> accessed 4 December 2020)

— — ‘Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences’ (UNHCR, September 2020) <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf> accessed 4 December 2020

United Nations Human Rights Council, ‘General Comment No 34: Article 19: Freedom of opinion and expression’ (HRC, 12 September 2011), UN Doc CCPRC/GC/34

— — ‘Report on the Israeli settlements in the Occupied Palestinian Territory, including East Jerusalem, and in the occupied Syrian Golan’ (30 January 2020) UN Doc A/HRC/43/67

United Nations Secretary-General, ‘Guide Note of the Secretary-General on Democracy’, UNSG <https://www.un.org/en/pdfs/FINAL%20Guidance%20Note%20on%20Democracy.pdf> accessed 3 June 2021

United Nations Security Council, Resolution 465 (1980) (S/RES/465, 1 March 1980); UNSC, Resolution 2334 (S/RES/2234, 23 December 2016)

United States of America

Central Intelligence Agency, Federal Bureau of Investigation, and National Security Agency, ‘Assessing Russian Activities and Intentions in Recent US Elections’ (6 January 2017) ***HTTPS://WWW.DNI.GOV/FILES/DOCUMENTS/ICA_2017_01.PDF***> accessed 10 May 2019

CISA, Joint Statement From Elections Infrastructure Government Coordinating Council & the Election Infrastructure Sector Coordinating Executive Committees’ (CISA, 12 November 2020) <https://www.cisa.gov/news/2020/11/12/joint-statement-elections-infrastructure-government-coordinating-council-election> accessed 28 January 2021

— — ‘COVID-19 Disinformation Toolkit’ (CISA.gov, 2020) <https://www.cisa.gov/sites/default/files/publications/SLTTCOVIDToolkit_FINAL_508.pdf> 30 April 2021

Department of Homeland Security, ‘Combatting Targeted Disinformation Campaigns: A whole-of-society issue’ (*DHS*, October 2019) <https://www.dhs.gov/sites/default/files/publications/ia/ia_combatting-targeted-disinformation-campaigns.pdf> accessed 3 March 2021

Department of State, *‘Country Report on Terrorism 2018’* (October 2019) <https://www.state.gov/wp-content/uploads/2019/11/Country-Reports-on-Terrorism-2018-FINAL.pdf> accessed 1 February 2019

Gohmert L, ‘Gohmert Introduces Bill That Removes Liability Protections for Social Media Groups that Use Algorithms to Hide, Promote, or Filter User Content’ (US Congressman Louie Gohmert Press Release, 20 December 2018) <https://gohmert.house.gov/news/documentsingle.aspx?DocumentID=398676> accessed 25 April 2019

McNamara R, ‘Report of the Office of the Secretary of Defense Vietnam Task Force’ (1967)

Oliva J, ‘Rule 6A-1.094124, Florida Administrative Code (F.A.C.), Required Instruction Reporting’ (Florida Department of Education, 14 June 2021) <https://info.fldoe.org/docushare/dsweb/Get/Document-8724/dps-2019-165.pdf> accessed 29 June 2021

US Congress, ‘Strengthening America’s Security in the Middle East Act of 2019 (ROS19020)’ (116th Cong 1st sess, 2019-2020)

US Congress, ‘Israel Anti-Boycott Act (HR 5595)’ (116th Cong, 2019-2020)

US House Committee on the Judiciary, ‘April 9: House Judiciary to Hold Hearing on Hate Crimes & White Nationalism’ (3 April 2019) <https://judiciary.house.gov/news/press-releases/april-9-house-judiciary-hold-hearing-hate-crimes-white-nationalism> accessed 14 May 2019

US Senate Committee on Foreign Relations, ‘Report on the International Covenant on Civil and Political Rights’, S Exec Rep 23, 1 (102d Sess. 1992), reprinted in (1992) ILM 645

— — ‘Report on the International Covenant on the Elimination of All Forms of Racial Discrimination’, S Exec Rep 103-29 (2d Sess 1994)

West K and others, ‘Enrolled House Bill No 1775’ <http://webserver1.lsb.state.ok.us/cf_pdf/2021-22%20ENR/hB/HB1775%20ENR.PDF> accessed 20 June 2021

White House, ‘Remarks by President Trump After Meeting with Congressional Leadership on Border Security’ (*Trump White House Archives,* 4 January 2019) <https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-meeting-congressional-leadership-border-security> accessed 25 January 2020

— — ‘Executive Order 13985: Advancing Racial Equity and Support for Underserved Communities Through the Federal Government’ (White House Briefing Room, 20 January 2021) <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/> accessed 20 June 2021

— — ‘FACT SHEET: National Strategy for Countering Domestic Terrorism (*White House Statements and Releases*, 15 June 2021) <https://www.whitehouse.gov/briefing-room/statements-releases/2021/06/15/fact-sheet-national-strategy-for-countering-domestic-terrorism/> accessed 15 July

Introduction

In the United States, individuals may declare support for terrorism, deny the holocaust, and express hatred against groups based on religion or race without fear of running afoul of the law.[1] In Europe, identical expression may result in substantial fines and lengthy prison sentences.[2] These examples highlight that while the notion that freedom of speech is a well-established and valuable right is uncontested in liberal democracies, significant disagreements exist with respect to how it is conceptualised in legal frameworks and the extent to which government restrictions on the message or content of expression are considered legitimate. These differences raise significant challenges that are the subject of longstanding debates. In the digital age, advances in technology have transformed these debates by, among other things, profoundly altering the ways in which people communicate with one another and how governments communicate with the public. As a result of these changes, governments and supranational bodies are enacting increasingly onerous measures directed to digital intermediaries - entities that provide online platforms for the expression of users - to implement and enforce aggressive content moderation practices.

In 1999, James Weinstein observed that while 'the goal of free speech doctrine can be easily stated: forbidding government from suppressing speech that must be permitted in a free and democratic society while allowing it to punish speech that causes harm that government may legitimately prevent…accomplishing this goal is not so easy'.[3] Accomplishing the goal of free speech doctrine in the digital age requires tackling important questions regarding contemporary efforts to regulate extreme speech, including whether such efforts are compatible with free speech principles and whether existing free speech doctrines, largely developed in the prior century, remain fit for purpose in the contemporary information ecosystem.

[1] See, e.g., *Matal v Tam,* 137 SCt 1744, 1764 (2017), in which the United States Supreme Court (USSC) opined that '[s]peech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express "the thought that we hate"'.

[2] For example, many European states, including France, the UK, and Germany criminalise terrorist-related expression.

[3] James Weinstein, *Hate Speech, Pornography, and Radical Attacks on Free Speech Doctrine* (Routledge 1999) 11.

Aims and Scope

The aim of this thesis is to critically examine the most challenging and significant contemporary free speech questions raised by three types of extreme speech - 'hate speech'[4], terrorist-related expression, and disinformation from state actors, through a comparative inquiry concerning the regulation of these types of speech in Europe and the United States. [5] The purpose of this inquiry is to offer an original contribution to the broader discourse concerning the appropriate limits on freedom of expression in liberal democracies in the digital age.

The scope of this inquiry reflects a few principal conclusions following extensive research. First, given the combination of the increasing prevalence and virulence of extreme expression online and increasingly aggressive efforts by states and supranational bodies to regulate online speech (often in ways that have transnational effects), research examining the appropriate limits of freedom of expression in the digital age is of particular value and topicality. Second, the most significant and challenging contemporary free speech questions relate to the online regulation of 'hate speech', terrorist-related expression, and disinformation from state actors.

Third, and related to the second point, the scope of this thesis reflects the conclusion that there is value in examining these three types of extreme speech in a single volume. Legal restrictions on 'hate speech' became commonplace during the post-World War II period, in which the international community identified a need for such restrictions and many European states enacted 'hate speech' bans.[6] Thus, there is a wealth of available evidence regarding the efficacy of such restrictions and the free speech implications flowing therefrom. At present, terrorist-related expression is the focus of increasingly aggressive measures directed at digital intermediaries, providing broader insights into the free speech implications of aggressive efforts to regulate the private actors that host so much of the expression in contemporary public

[4] With respect to the term 'hate speech', this thesis adopts the approach of Peter Molnar, who observes that 'the colloquial expression "hate speech" seems to presuppose that a government can define with legal precision the categories of expression that warrant regulation as "hate speech"'. Like Molnar, I question this implicit assumption and, for this reason, use 'hate speech' only in quotation marks. In so doing, I am not refuting the assumption that some expression may reasonably be classified as 'hate speech' but, rather, am emphasising that the definition of this term remains normatively and legally contested. See Peter Molnar, 'Towards Better Law and Policy Against "Hate Speech" – The "Clear and Present Danger" Test in Hungary', in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy'* (OUP 2009) 237.

[5] Chapters 3, 4, and 5, respectively.

[6] Richard Moon, *Putting Faith in Hate* (CUP 2018) 23 - 24.

discourse.[7] Disinformation, particularly from state actors, is an emerging and increasingly dangerous challenge in the digital age, which is beginning to attract the focus of regulators at the national and supranational levels. The lessons from efforts to regulate 'hate speech' and terrorist-related expression, including but not limited to the importance of precise definitions in regulatory frameworks and the potential implications of intermediary liability, will be instructive in developing and evaluating frameworks directed at regulating disinformation. Thus, engaging with these types of extreme speech within a broader inquiry into the appropriate limits on freedom of expression in the digital age highlights the unique harms and regulatory challenges flowing from each type of speech while illuminating overlapping and interrelated issues from which both broader and more nuanced themes, lessons, and connections emerge.

Finally, the scope of this inquiry reflects the position that situating such an inquiry within a broader comparative analysis of the approaches of Europe and the United States provides for richer and more nuanced observations than result from inquiries focusing on a single jurisdiction. The decision to compare the approaches of Europe and the United States reflects the latter's position as anomalous among liberal democracies for its expansive free speech protections and the emergence of post-World War II European democracies within an integrated supranational system that has made Europe a 'centre of gravity - a crucial political and culture counter-weight to the [United States]'.[8] Additionally, Europe and the United States are key players in the transnational discourse concerning whether digital intermediaries, in particular Facebook, Twitter, and other large platforms, should be required to do more to identity and remove extreme content.[9] While the United States is, for the moment, maintaining a statutory framework that provides expansive immunity to digital intermediaries for the expression of third party users, Europe has introduced compulsory frameworks that impose onerous obligations on digital intermediaries, the largest and most influential of which are based in the United States and led by Americans. [10]

[7] Regulations on terrorist-related expression became common-place at the national and supranational levels following the 11 September terrorist attacks in New York.
[8] Eric Heinze, 'Wild-West Cowboys versus Cheese-Eating Surrender Monkeys: Some Problems in Comparative Approaches to Hate Speech' in Hare and Weinstein (n 4) 186.
[9] See Joris van Hoboken and Daphne Keller, 'Design Principles for Intermediary Liability Laws' (8 October 2019) Transatlantic Working Papers Series, 3 - 4 <https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/Intermediary_Liability_TWG_van_Hoboken_Oct_2019.pdf> accessed 10 July 2021.
[10] The headquarters of Facebook, Twitter, Google, and YouTube are located in California and the CEO of each company is American.

This inquiry is informed by what Robert Post describes as 'conceptual precision'. By conceptual precision, Post refers to an appreciation for how the same opinions and ideas expressed in different contexts garner varied constitutional protection within a particular jurisdiction.[11] For example, while the United States is rightly regarded as an outlier amongst liberal democracies for its rigorous protection of extreme speech, it is important to appreciate that such protection applies almost exclusively in the domain of public discourse, and that there is extensive regulation of such expression in other areas of the American legal framework that raise no First Amendment concerns, including employment law.[12] This thesis also adopts Post's definition of the term public discourse, which includes 'all communicative processes deemed necessary for the formation of public opinion'.[13] While Post's observations concern the United States, they apply with equal force in Europe, where speech concerning matters of public debate receives the highest degree of protection from government interference.[14] With respect to speech falling outside the scope of public discourse, such as the workplace and courtrooms, restrictions on expression are commonplace in both Europe and the United States, do not raise the ire of free speech advocates on either side of the Atlantic, and lie outside the scope of this thesis.[15]

Conceptual precision requires adopting and applying clear definitions of the primary terminology used in this thesis. Throughout this thesis, I refer to the American and European approaches to freedom of expression. By American approach, I primarily mean the approach of the United States Supreme Court (USSC) though, reference to, and analysis of, decisions of lower federal courts are provided when instructive in fleshing out particular principles and concepts. Legislation and legislative proposals are also considered when relevant to discussions of emerging trends and regulatory efforts at the state and federal levels. By European approach, I primarily mean the approach of the European Court of Human Rights (ECtHR), which is tasked with securing the rights and freedoms contained in the European

[11] See Robert Post, 'Interview with Robert Post' in Michael Herz and Peter Molnar (eds*), The Content and Context of Hate Speech: Rethinking Regulation and Responses* (CUP 2012) 12 - 13.
[12] ibid.
[13] Robert Post, 'Participatory Democracy and Free Speech' (2011) 97(3) Va L Rev 477, 486.
See also *Sorrell v IMS Health, Inc*, 564 US 552, 582 (2011), in which the USSC held that 'the First Amendment imposes tight constraints upon government efforts to restrict, *e.g.,* "core" political speech, while imposing looser constraints when the government seeks to restrict, *e.g.,* commercial speech, the speech of its own employees, or the regulation-related speech of a firm subject to a traditional regulatory program'.
[14] See, e.g. *Dichand and others v Austria*, App no 29271/95 (ECtHR, 26 February 2002), para 154, in which the European Court of Human Rights (ECtHR) held that '[t]here is little scope … for restrictions on political speech or debates on questions of public interest'.
[15] See 'Interview with Robert Post' in Herz and Molnar (n 11) 11 - 12.

Convention on Human Rights (ECHR or Convention) - the most comprehensive and developed system for supranational human rights protection in the world[16] - and providing an effective remedy before a national authority for alleged violations in the 47 Council of Europe Member States (Member States).[17]

The emphasis on the ECtHR reflects its status as 'the leading instrument for the protection of fundamental rights in Europe'.[18] It further accounts for the EU's recognition that the right to freedom of expression in the Charter of Fundamental Rights of the European Union (EU Charter) is given the same meaning and scope as Article 10 of the ECHR[19] and reflects the EU's legal obligation under Article 6(2) of the Treaty of Lisbon to accede to the ECHR.[20] The purpose of the EU's accession to the ECHR is to 'contribute to the creation of a single European legal space, achieving a coherent framework for human rights protection throughout Europe'.[21] Formal negotiations between EU Member States and the European Commission resumed in September of 2020 following stalled negotiations in 2014.[22] This same month, the Secretary General of the Council of Europe and the EU Commission's Vice President for Values and Transparency issued a joint statement asserting that accession 'will help to guarantee coherence and consistency between EU law and the Convention system' and 'will also ensure that the EU is subjected to the same international oversight on human rights as its 27 member states and 20 other Council of Europe countries which are not members of the EU'.[23]

When relevant to specific inquiries, the approach of the EU is also considered. While the EU and the Council of Europe are separate entities, the EU's actions impact on the scope and application of freedom of expression for Member States, all of which are also EU Member

[16] Ivan Hare, 'Extreme Speech under International and Regional Human Rights Standards', in Hare and Weinstein (n 4) 66. For a comprehensive overview of the ECHR, see Janneke Gerards, *General Principles of the European Convention on Human Rights* (CUP 2019).
[17] 'European Union accessed to the European Convention on Human Rights - Questions and Answers' (*Council of Europe*) <https://www.coe.int/en/web/portal/eu-accession-echr-questions-and-answers> accessed 2 May 2021.
[18] Ottavio Marzocchi, 'The protection of Article 2 TFEU values in the EU', European Parliament Fact Sheet (April 2021) <https://www.europarl.europa.eu/factsheets/en/sheet/146/the-protection-of-article-2-teu-values-in-the-eu> accessed 2 July 2021.
[19] Charter of the Fundamental Rights of the European Union [2012] OJ C326/L391. Article 52(3) instructs that '[i]n so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention'.
[20] Article 6(2) of the Consolidated Version of the Treaty on the European Union [2008] OJ C115/13.
[21] Council of Europe 'Questions and Answers' (n 17).
[22] ibid.
[23] 'The EU's accession to the European Convention on Human Rights: Joint statement on behalf of the Council of Europe and the European Commission (*Commission Press Corner*, 29 September 2020) <https://ec.europa.eu/commission/presscorner/detail/en/statement_20_1748> accessed 13 March 2021.

States. For example, the EU recently enacted legislation in the area of online-terrorist related expression. Because EU legislation is binding on EU Member States, the EU's decision to legislate in this area greatly impacts the regulation of this type of extreme speech in much of Europe (and potentially far beyond). Thus, EU legislation is examined in Chapter 4, which discusses terrorist-related expression. Additionally, while it is beyond the scope of this thesis to provide an exhaustive account of the approaches of all EU Member States, the approaches of particular States are examined when instructive to the examinations undertaken in particular chapters. For example, Chapter 3 discusses Germany's online hate speech law, which was the first national legislation to hold social media platforms responsible for combating online 'hate speech' and is serving as a template for similar laws in other jurisdictions. The lessons from domestic disinformation campaigns from the Hungarian government, arguably the worst offender for domestic disinformation in Europe, are considered in Chapter 6, which examines disinformation from state actors.

Additionally, this thesis uses the terms 'constitutional rights' and 'constitutional frameworks' broadly to include rights and frameworks that derive from national constitutions as well as international treaties notwithstanding the fact that the latter, as discussed below, are distinct from constitutions.[24] The term 'government' refers to executive and legislative branches, collectively, unless expressly stated otherwise. It is also important to be clear at the outset with respect to what lies outside the scope of this thesis. While the regulation of extreme speech in liberal democracies raises important ethical and moral issues, in particular with regard to the social responsibilities of digital intermediaries, these are not of concern here. Additionally, while the historical development of European and American constitutional frameworks is relevant to a comprehensive understanding of the European and American approaches to free speech, these developments are discussed only when instructive for the discussions and explorations undertaken in this thesis. Further, there is voluminous literature on the extensive debates regarding the connection between 'hate speech' and the types of harm that may be considered sufficient to justify restrictive measures; this thesis does not attempt to add to the literature in this area.[25] However, it is important to locate its stance within these debates.

[24] This approach is similar to that of Kai Möller in *The Global Model of Constitutional Rights* (OUP 2015).

[25] Many scholars have waded into debates regarding the harm caused by 'hate speech'. Among the most notable of these include critical race theorists Richard Delgado and Mari Matsuda (see Mari Matsuda and others (eds), *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Westview Press (1993)), Jeremy Waldron (see Jeremy Waldron, *The Harm in Hate Speech* (Harvard University Press 2012)); and Nadine Strossen (see Nadine Strossen, *HATE: Why We Should Resist it With Free Speech, Not Censorship* (OUP 2018)).

Accordingly, this thesis accepts the argument that 'hate speech' causes less tangible forms of harm than physical violence that negatively impact public discourse and foment societal hate against vulnerable and marginalised groups. The focus of this thesis is on legal efforts to ameliorate the different types of harm flowing from 'hate speech, as well as terrorist-related expression and disinformation from state actors. It interrogates the efficacy of such efforts in relation to government objectives as well as the potential implications to the protection of freedom of expression.

In approaching this inquiry, I am mindful of James Weinstein's observation that constitutional law scholarship in general and free speech scholarship, in particular, are 'marked by an unhelpful tendency to confuse the descriptive – what the law is – with the normative – what the law should be'.[26] Accordingly, the examinations undertaken in this thesis carefully distinguish between descriptive and normative elements. Finally, it is important to highlight that the examinations undertaken and observations made in this thesis reflect events at a particular point in time. The discourse on free speech in the digital age is constantly evolving as advancements in technology create new pathways for communication as well as new challenges and opportunities for regulation. This thesis is current as of 27 July 2021. Relevant events occurring after this date, of which there are sure to be many, will be the subject of future research projects and may alter the recommendations and conclusions reached in this volume.

Methodology

The methodology used in this thesis is that of comparative constitutional law (CCL). CCL scholarship seeks to develop a detailed understanding of how people living in different cultural, social and political contexts deal with common constitutional questions that are assumed to be common to the majority of modern political systems.[27] The most basic definition of CCL is the 'study of constitutional systems or their various components across time and place, with an aim of generating some kind of analytical yield by the act of comparison'.[28] The epistemological focus of CCL is often on the internal logic, hierarchy and interpretive

See also Ishani Maitra and Mary Kate McGowan, 'On Racist Hate Speech and the Scope of a Free Speech Principle' (2007) 23(2) CJLJ 343. Waldron's theory of the relationship between harm and 'hate speech' is discussed in Chapter 2.

[26] Weinstein (n 3) 10 - 12.

[27] Ran Hirschl, 'Comparative Methodologies', in Roger Masterman and Robert Schütze (eds), *The Cambridge Companion to Comparative Constitutional Law* (CUP 2019) 24 - 25.

[28] ibid 16.

coherence of constitutional law as an autonomous legal system.[29] Such study may be 'descriptive, taxonomical, hermeneutic, conceptual, normative, explanatory or any combination of these and other types of scholarly inquiry'.[30]

CCL is used in this thesis for the purpose of identifying the 'best' or 'most suitable' norms and practices concerning the question of the appropriate limits on freedom of expression in liberal democracies in the digital age.[31] While the challenges confronting the United States and Europe relating to the causes and consequences of extreme speech are similar, the approaches vary in significant and meaningful ways. For example, while Europe increasingly permits content-based restrictions on speech directed to denigrating marginalised and vulnerable groups, the United States holds steadfast to the principle that societal problems cannot be addressed by restricting expression based on viewpoint and that more harm than good results from such measures. While it may be temping, given these differences, to adopt a purely contextual approach to extreme speech that considers a broad range of constitutional responses and the impact on free speech rights, globalisation and the transnational nature of online expression make a purely contextual approach inadequate.[32]

This thesis is based on the premise of the value of a comparative inquiry grounded in the doctrinal differences of the USSC and the ECtHR in light of the important similarities that make this type of inquiry instructive. It is important, however, to acknowledge at the outset that the selection of the ECHR and the ECtHR as comparators with the United States Constitution and the USSC may raise questions given that the ECHR is an international treaty rather than a constitution. This inquiry builds on the work of other scholars, including Ian Cram, Kai Möller, and Eric Heinze, who have undertaken CCL examinations of this type while acknowledging the ways in which this difference manifests in the interpretation and application of the First Amendment and Article 10 by the USSC and the ECtHR, respectively.[33]

---

[29] ibid 13 - 14.
[30] ibid 16.
[31] See ibid 17.
[32] See Michel Rosenfeld, 'Hate Speech in Constitutional Jurisprudence' in Herz and Molnar (n 11) 242 - 243.
[33] See, e.g., Ian Cram, 'Coercing Communities or Promoting Civilised Discourse? Funeral Protests and Comparative Hate Speech Jurisprudence' (2012) 12 HRL Rev 455; Möller (n 24); and Erik Heinze (n 8). See also Kanstantsin Dzehtsiarou and Conor O'Mahony, 'Evolutive Interpretation of Rights Provisions: A Comparison of the European Court of Human Rights and the U.S. Supreme Court' (2013) 44 Colum Hum Rts L Rev 309; Jeffrey A Brauch, 'The Dangerous Search for and Elusive Consensus: What the Supreme Court Should Learn from the European Court of Human Rights' (2009) 52 Howard LJ 277.

For example, in the context of a comparative evaluation of freedom of expression, Ian Cram highlights that 'it is important moreover to note the clear structural differences which exist between the First Amendment and Article 10 of the [ECtHR]'.[34] These structural differences reflect, among other things, the fact that unlike constitutions, international human rights treaties are voluntary agreements between sovereign states that may accede or withdraw at their discretion.[35] Thus, an international court or supervisory body has no jurisdiction to hear complaints from individuals within a state without that state's consent. This is true of the ECHR, notwithstanding that it provides for direct access to courts and that its jurisdiction is compulsory and binding on Member States.[36] Moreover, while the ECtHR has repeatedly characterized the ECHR 'as a constitutional instrument of European public order (ordre public)',[37] it belongs to a different type of legal order than national courts in Europe and does not exercise its judicial function in the domestic legal order of Member States.[38] Kanstantsin Dzehtsiarou describes the institutional position of the Court as a 'legitimacy deficit' stemming from its status as an international court that must contend with the sovereignty of Member States and from its authority to 'call into question the decisions of democratically elected governments based on vaguely defined human rights norms'.[39]

Additionally, unlike the highest court in a national legal framework, the ECtHR is responsible for managing the interests of Member States and, in so doing, confronts divisive social and cultural issues within a broader context in which the cooperation of these States is required to secure the execution of its judgments.[40] Because the ECtHR is not an appellate court, it has no authority to reverse the judgment of the national court.[41] Rather, it may declare the national court's decision (and the relevant domestic law) in violation of the Convention. The domestic

[34] ibid (Cram) 456 - 457. Cram notes that '[f]or example, the US Supreme Court has had to deal with questions about what precisely is protected by the notion of 'speech' in the first place, whilst the Convention's more inclusive notion of 'expression' has shifted attention onto states' justifications for interference with expression'.
[35] See George Letsas, 'The ECHR as a living instrument: its meaning and legitimacy,' in Andreas Føllesdal, Birgit Peters and Geir Ulfstein (eds), *Constituting Europe: The European Court of Human Rights in a National, European and Global Context* (CUP 2013). However, as noted above, the EU's accession became a legal requirement by way of the Lisbon Treaty.
[36] ibid 309.
[37] See, e.g., *Al-Skeini and Others v the United Kingdom*, [GC] App no 55721/07 (2011) para 141.
[38] Geir Ulfstein, 'The European Court of Human Rights as a Constitutional Court?' (2014) PluriCourts Research Paper No. 14-08.
[39] Kanstantsin Dzehtsiarou, *European Consensus and the Legitimacy of the European Court of Human Rights* (CUP 2015) 142 - 143.
[40] See Dzehtsiarou and O'Mahony (n 33). See also Alain Zysset, 'Searching for the Legitimacy of the European Court of Human Rights: The Neglected Role of "Democratic Society"' (2016) 5(1) GlobCon 16, in which he discusses the recent trend of politicians, academics, and national judges questioning the ECtHR's authority and role as the final interpreter of the ECHR.
[41] Dieter Grimm, 'Freedom of Speech in a Globalized World' in Hare and Weinstein (n 2) 21.

judgment remains in place, though, the Member States is under a duty to act compatibly with the Convention.[42] Thus, the ECtHR must contend with the tension arising from its task of protecting human rights with its obligation to respect the democratic sovereignty of Member States.[43] This reality influences how the ECtHR interprets and applies Convention rights.[44]

To navigate these tensions, the ECtHR has adopted interrelated interpretive mechanisms for adjudicating disputes over Member States' restrictions on ECHR rights. Subsidiarity - the principle that it is 'not the Court's task to substitute its own assessment of the facts for that of the domestic courts and it is for the latter to establish the facts on the basis of the evidence before them'[45] - reflects the ECHR's supplementary role to national constitutions with respect to safeguarding human rights.[46] The ECtHR uses subsidiarity to justify a wide margin of appreciation - which refers to the operating space that the national authorities enjoy in fulfilling their obligations under the Convention - to Member States in the absence of a European consensus.[47] The doctrine of European consensus refers to the level of uniformity present in the legal frameworks of Member States on a particular topic.[48] The narrower the degree of consensus, the greater the margin of appreciation left to Member States and vice/versa.[49]

The margin of appreciation doctrine plays an important and contested role in the ECtHR's adjudication of ECHR rights, highlighting the 'mediating role' of the principle of subsidiarity 'in finding an appropriate equilibrium between national constitutional protection systems on the one hand, and regional or universal systems on the other'.[50] Under the doctrine, freedoms

[42] ibid.
[43] Andreas Føllesdal, 'Subsidiarity and International Human-Rights Courts: Respecting Self-Governance and Protecting Human Rights - or Neither?' (2016) 79 LCP 147, 148.
[44] ibid.
[45] *Austin and others v United Kingdom*, App nos 39692/09, 40713/09 and 41008/09 (ECtHR, 15 March 2012). See also Alastair Mowbray, 'Subsidiarity and the European Convention on Human Rights' (2015) 15 HRL Rev 313.
[46] See Yutaka Arai-Takahashi, 'The margin of appreciation: A theoretical analysis of Strasbourg's variable geometry' in Føllesdal, Peters and Ulfstein (n 35). See also Steven Greer, Janneke Gerards and Rose Slowe, *Human Rights in the Council of Europe and the European Union* (CUP 2018); Andrew Legg, *The Margin of Appreciation in International Human Rights Law Deference and Proportionality* (OUP 2012).
[47] See 'Interpretive Mechanisms of ECHR case-law: the concept of European Consensus', Council of Europe <https://www.coe.int/en/web/help/article-echr-case-law> accessed 9 July 2021.
[48] See Eyal Benvenisti, 'Margin of Appreciation, Consensus, and Universal Standards, 31 NYUJ Intl L & Pol 843 (1999) 851-52.
[49] ibid.
[50] Arai-Takahashi (n 46) 90-91. For in-depth examinations of the margin of appreciation doctrine, see Steven Greer, 'The Margin of Appreciation: Interpretation and Discretion under the European Convention on Human Rights' (Council of Europe Publishing 2000); Ronald St John Mcdonald, 'The Margin of Appreciation', in Ronald St John Mcdonald, Frantz Matscher and Herbert Petzold (eds), *The European System for the Protection*

considered fundamental in any democratic society, including freedom of expression, may enjoy varying protection across Member States.[51] It takes into account the different legal systems operating in Member States, as well as the different approaches to addressing issues that states with the same legal system may adopt. In many contexts, the width of the margin of appreciation will greatly impact the degree to which the Court scrutinises alleged violations of ECHR rights.[52] Thus, the scope of the margin of appreciation accorded to a Member State in a given case can be crucial to the ECtHR's ultimate determination of whether an interference violated the Convention.[53]

The significance of the institutional position of the ECtHR is illustrated by recent developments in Russia. In March of 2021, an amendment to the Russian Constitution, adopted by the Russian Parliament and signed into law by President Vladmir Putin, added a provision stipulating that international agreements, treaties, and decisions by international bodies are valid only to the extent that they do not contradict the Russian Constitution.[54] This amendment empowers the Russian Constitutional Court to declare as non-executable any international decisions, including those of the ECtHR, that contradict the Russian Constitution.[55] The Venice Commission issued an opinion on the draft amendment as part of the fact-finding for a report on the implementation of the judgments of the ECtHR in which it highlighted, among other things, that the 'judgment of an international court implie[s] a deliberate balance between international jurisdiction and national sovereignty. Its enforcement therefore calls for a different type of procedure from that applicable to national proceedings, involving, among

of Human Rights (Martinus Nijhoff 1993). For a comprehensive and critical examination of the application of the margin of appreciation doctrine in the jurisprudence of the ECtHR, see George Letsas, 'Two Concepts of the Margin of Appreciation' (2006) 26(4) OJLS 705; Dean Spielman, 'Allowing the Right Margin in the European Court of Human Rights and the National Margin of Appreciation Doctrine: Waiver or Subsidiarity of European Review?' (2012) 14 CYELS 381, updated at <http://www.echr.coe.int/ Documents/Speech_20140113_Heidelberg_ENG.pdf>).

[51] Ian Cram, 'The Danish Cartoons, Offensive Expression, and Democratic Legitimacy' in Hare and Weinstein (n 4) 317.

[52] See Philippe Yves Kuhn, 'Reforming the Approach to Racial and Religious Hate Speech under Article 10 of the European Convention on Human Rights' (2019) 19 HRL Rev 119, 144 - 145. See also Helen Fenwick and Daniel Fenwick, '"East"/"West" Divisions in Council of Europe States on Treatment of Sexual Minorities: The Response of Strasbourg Court and the Role of Consensus Analysis' (2019) 3 EHRLR 247.

[53] Dominic McGoldrick, 'A Defence of the Margin of Appreciation and an Argument for its Application by the Human Rights Committee' (2016) 65 ICLQ 21, 24.

[54] See Isabelle Jeffries, 'Russia's Constitutional Amendment from an International Law Perspective' (*Public International Law & Policy Group*, 1 March 2021).

[55] This is illustrative of Russia's broader compliance problem in the context of the European human rights system. See Lauri Mälksoo, 'Russia's Constitutional Court Defies the European Court of Human Rights' (2016) 12(2) Eur Const Law Rev 377.

other things, dialogue and cooperation'.[56] This may explain, in part, why enforcement of ECtHR judgments by Member States is a major challenge for the Council of Europe.[57] In 2018, more than half of the nearly 75,000 judgments issued by the ECtHR since its inception remained unenforced.[58]

In contrast to the ECtHR, the USSC is the highest national court within a federalist system in which its legitimacy and supremacy - as well as the legitimacy and supremacy of the federal constitution - are neither voluntary nor in doubt. As such, its decisions are directly effective and enforceable against individual states.[59] The Supremacy Clause of the United States Constitution instructs that the Constitution 'and the law of the United States which shall be made, under the authority of the United States, shall be supreme law of the land; and judges in every state shall be bound thereby, anything in the Constitution or laws of any State to the contrary notwithstanding'. [60] In the seminal case of *Marbury v. Madison* in 1803, the USSC held that the Constitution is 'the *supreme* law of the land' and that '[i]t is emphatically the province and duty of the judicial department to say what the law is' because '[t]hose who apply the rule to particular cases, must of necessity expound and interpret that rule'.[61] This reflects the institutional position and inherent authority of the USSC, the judgments of which all other courts, including the supreme courts of individual states, must respect.[62]

---

[56] 'Russian Federation Opinion on the Draft Amendments to the Constitution (As Signed by the President of the Russian Federation on 14 March 2020) Related to the Execution in the Russian Federal of Decisions by the European Court of Human Rights (18 June 2020) Opinion No 981/2020 (*Venice Commission*, 2020) 15 (quoting Seminar background paper: Implementation of the judgments of the European Court of Human Rights: a shared judicial responsibility?) However, the Venice Commission further notes that '[t]he Russian Federation has made the political decision to join the Council of Europe', 'has committed itself to executing the judgements of the Court', and that 'there is no choice to execute or not execute the Strasbourg Court judgement' as 'the judgements of the ECtHR' are binding. ibid 16.

[57] Veronika Fikfak, 'Changing State Behaviour: Damages before the European Court of Human Rights' (2018) 29(4) EJIL 1091, 1092 - 1093. See also Fiona De Londras and Kanstantsin Dzehtsiarou, 'Mission Impossible? Non-Execution Through Infringement Proceedings in the European Court of Human Rights (2017) 66 ICLQ 467.

[58] See Council of Europe, 'Supervision of the Execution of Judgments and Decisions of the European Court of Human Rights 2017: 11th Annual Report of the Committee of Ministers' (March 2018) 7.

[59] *Marbury v Madison*, 5 US 137, 180 (1803).

[60] US Const Article VI.

[61] *Marbury* (n 59) (emphasis added).

[62] See *Nitro-Lift Technologies, LLC v Howard*, 568 U.S. 17, 21 (2012) (internal citations omitted) (opining that '[t]he state court insisted that its "[own] jurisprudence controls this issue" and permits review of a "contract submitted to arbitration where one party assert [s] that the underlying agreement [is] void and unenforceable." But the Oklahoma Supreme Court must abide by the [Federal Arbitration Act], which is the supreme Law of the Land, and by the opinions of this Court interpreting that law. It is this Court's responsibility to say what a statute means, and once the Court has spoken, it is the duty of other courts to respect that understanding of the governing rule of law. Our cases hold that the FAA forecloses precisely this type of judicial hostility towards arbitration'.).

Notwithstanding these differences, there are important similarities that make the type of inquiry undertaken in this thesis instructive. George Letsas notes that the similarities between the ECHR and national constitutions include 'lists of abstract fundamental rights that individuals have against government' and that 'both create judicial institutions with the power to review whether individuals' *legal* rights have been violated'.[63] He also persuasively argues that the ECHR is 'best seen as enshrining human rights that are both legal and liberal: they are founded upon liberal egalitarian principles that impose conditions on the legitimate use of coercion by member states against persons within their jurisdiction'.[64] As a result, 'the normative role of the Convention rights is therefore no different to that of domestic constitutional rights within a liberal democracy'. Additionally, Letsas observes that '[j]ust like a constitutional court, the [ECtHR] has the final authority to rule on whether a state (through its statutory provisions, case law, or executive acts) violates abstract moral principles' and what the ECtHR adjudicates on 'is inevitably an abstract issue of principle which it then must apply to all Europeans'.[65] Kai Möller highlights that the ECtHR performs a review function very similar to that of the highest domestic courts of states – including the United State Supreme Court (USSC) - making a comparative analysis both appropriate and instructive.[66]

Finally, it is worth noting Mark Tushnet's advice that CCL scholars be cognisant of the differences in doctrinal structures in different jurisdictions as the precise doctrinal formulation of a particular constitutional standard 'might be quite consequential'.[67] By way of example, he notes that while American constitutional scholars appreciate the difference between the 'clear and present danger' test, common in many legal frameworks, and the less well-known *Brandenburg* test that is unique to the United States, scholars from other jurisdictions might not appreciate such doctrinal nuances.[68] As a former attorney for the American Civil Liberties Union with an LLM in Public International Law and experience teaching public law at the university level in Europe, I am uniquely positioned to undertake a CCL inquiry of this type

[63] George Letsas 'ECHR as a living instrument' (n 35) 107.
[64] George Letsas, *A Theory of Interpretation of the European Convention on Human Rights* (OUP 2009) 9 - 10.
[65] ibid. See also Michel Rosenfeld, 'Comparing constitutional review by the European Court of Justice and the U.S. Supreme Court' (2006) 4(4) ICON 618, 621, in which Rosenfeld opines that the ECtHR, 'while a transnational court interpreting and applying ECHR treaty-based rights, engages substantively in something very much akin to the adjudication of constitutional rights'.
[66] Kai Möller (n 24) 16 - 17.
[67] Mark Tushnet, 'Comparative Constitutional Law' in Mathias Reimann and Reinhard Zimmeran (eds), *The Oxford Handbook on Comparative Law* (OUP 2019) 23.
[68] ibid. See *Brandenburg v Ohio*, 395 US 444 (1969) (holding that the government may only proscribe inflammatory speech intended to advocate illegal action if such speech is directed to inciting or producing imminent lawless action and is likely to produce such action.).

and to engage in critical and nuanced doctrinal examinations within European and American legal frameworks.

Original Contribution to the Broader Discourse

Before starting this inquiry, it is important to explain how this volume provides an original contribution to the broader discourse on the regulation of freedom of expression in liberal democracies in the digital age. As discussed above, notwithstanding the notable differences between the ECHR and the United States Constitution, many scholars engage in CCL inquiries of this type in order to offer insights into broader debates. In the area of freedom of expression, many doctrinal comparisons focus on a particular type of extreme speech, such as 'hate speech' or disinformation.[69] I am unaware of another work that examines in a single volume the unique harms and regulatory challenges of 'hate speech', terrorist-related expression, and disinformation in the contemporary digital context. This approach provides insights that do not emerge from inquiries focusing on a single type of speech, including the relative dangers of each type of speech as well as the relative efficacy of regulatory approaches directed to particular types of expression. Thus, while it is common for entire volumes to address a particular type of extreme speech, the decision to examine three types of extreme speech in this thesis is both deliberate and strategic.

I am also unaware of another work that undertakes a comparison of the European and American approaches to 'hate speech', terrorist-related expression, and disinformation at such a granular level, rigorously examining the nuances of relevant doctrines and philosophical principles and delving into both inter and intra-jurisdictional debates regarding the proper scope and application of free speech rights. In so doing, this inquiry exposes unsettled and contested elements in each approach as well as the ways in which existing doctrines may warrant re-examining - or abandoning altogether - given the ways in which the digital age has permanently altered how governments and citizens communicate in public discourse. It also offers recommendations for how to address the challenges posed by extreme speech in the digital age in light of the unique and varying harms caused by 'hate speech', terrorist-related expression,

[69] See, e.g., Roger Kiska, 'Hate Speech: A Comparison between the European Court of Human Rights and the United States Supreme Court Jurisprudence' (2012) 25 Regent U L Rev 107; Oreste Pollicino and Elettra Bietti, 'Truth and Deception across the Atlantic: A Roadmap of Disinformation in the US and Europe' (2019) 11 Italian J Pub L 43; Heinze (n 8); Rosenfeld 'Comparing Constitutional Review' (n 65); Robert Post, 'Legitimacy and Hate Speech' (2017) 32 Const Comment 651.

and disinformation, taking account of the lessons to be learned from recent regulatory efforts in Europe and the United States.

Structure and Outline of Chapters

This thesis is structured around one primary question: what are the appropriate limits of freedom of expression in the digital age? This question is addressed in two parts. The first two chapters, which comprise Part I, provide the doctrinal and theoretical scaffolding, respectively, for the examinations and arguments provided in Part II. In general terms legal doctrine tells us what speech is, while philosophy tells us why it is deserving of protection from interference by the state. In the context of free speech, Eric Barendt observes that '[i]t is impossible to draw a sharp line between legal and philosophical argument'.[70] Thus, deriving the maximum benefit from a comparative inquiry of this nature requires a comprehensive understanding of the doctrinal and philosophical foundations of freedom of expression in European and American constitutional frameworks. Additionally, the examinations undertaken in Part I lead to a richer and more contextualised understanding of each approach by, among other things, situating the relevant instruments and conceptions of rights in their historical contexts. The United States Constitution and the ECHR may be understood as responses to historical events at different times in difference places. These differences are important for contextualising the doctrinal and philosophical groundings of the American and European approaches to speech regulation.

Chapter 1 outlines the key doctrinal principles that inform the protection afforded to freedom of expression by the USSC and the ECtHR. It considers three threshold questions: what constitutes speech within constitutional frameworks; what classes of speech fall outside of such protections; and who is obligated to respect free speech rights. This inquiry reveals that while freedom of expression is an essential part of European and American constitutional frameworks, the doctrinal foundations of free speech in Europe and the United States differ in meaningful ways that are often overlooked in the broader discourse surrounding freedom of expression. Additionally, the broader discourse tends to gloss over the intra-jurisdictional contestation and discord that illuminate the inherent challenges of delineating exceptions to free speech within frameworks premised on expansive protection of expression in public discourse.

[70] Eric Barendt, *Freedom of Speech* (OUP 2005) 2.

Chapter 2 provides the philosophical underpinning for this thesis by examining the relationship between European and American free speech doctrine and legal philosophy in the area of extreme speech. It is important to note here that the primary aim of this chapter is to provide the necessary philosophical foundation for Part II by examining the ways in which the work of legal philosophers enriches our understanding of the relevant doctrinal divergences in European and American constitutional frameworks. The following questions inform this chapter: what do the philosophical underpinnings of the right to freedom of expression in Europe and the United States tell us about the ways in which Europeans and Americans view the role of government in regulating behaviour? What background assumptions about citizenship do laws governing freedom of expression implicate? Answers to these questions are sought by examining the ways in which common philosophical justifications for the protection of freedom of aid in interpreting and appreciating the marked differences in the American and European approaches to the regulation of extreme speech. This examination highlights how the ways in which the ECtHR and USSC interpret the scope of free speech rights are informed by the broader philosophical traditions underpinning these systems, which reflect deeply rooted philosophical disagreements that may be explained, in part, by the socio-historical and cultural climate in which those frameworks developed.

The second part of the thesis, Part II, relies on the doctrinal and philosophical scaffolding laid in Part I in examining the challenges raised by three types of extreme speech in the digital age. It is worth noting here that while the primary aim of Part I is to provide the doctrinal and theoretical grounding for Part II, the examinations undertaken Chapters 1 and 2 are not merely descriptive. For example, Chapter 6 builds off of the discussions in Chapter 2 to challenge the philosophical underpinnings of the American approach by arguing that theories of democratic legitimacy warrant re-examining the digital age, and that certain doctrines outlined in Chapter 1 are no longer fit for purpose.

The focus of **Chapter 3** is 'hate speech'. While the regulation of 'hate speech' in liberal democracies is not a recent phenomenon, given increasing levels of hate within and across national borders, the prevalence of 'hate speech' related offences in national and supranational frameworks, and increasingly aggressive efforts to regulate the online platforms on which so much contemporary expression takes place, it is more important than ever to examine the free speech implications of government efforts to regulate expressions of hate in public discourse.

This chapter argues that contemporary regulatory efforts are grounded in frameworks that are burdened by critical unanswered questions, including the normative justifications for regulating this type of expression in the first place and the propriety of using human rights penality in this context. While it is imperative to address these underlying questions in order to gain a full appreciation of the free speech implications of online efforts to regulate 'hate speech', the contemporary discourse often glosses over these issues, instead emphasising the importance of removing 'hate speech' from online platforms. Examining and offering insights into these fundamental questions is a necessary part of debates over the regulation of online 'hate speech', and ignoring such questions risks applying fundamentally flawed frameworks to broad swaths of online expression. Thus, Chapter 3 focuses on the ways in which regulations on 'hate speech', both online and offline, are conceptualised and applied in contemporary European and American legal frameworks and the extent to which recent developments in Europe and the United States may inform broader debates regarding the appropriate role of government in regulating extreme expression online.

Chapter 4 engages with recent efforts to regulate online terrorist-related expression and the appropriate role of digital intermediaries in such efforts. While there is overlap between 'hate speech' and terrorist-related speech in terms of content, this thesis treats these as distinct categories of expression owing to the fact that the European approach often distinguishes between terrorist-related expression and other forms of virulent speech. This chapter interrogates the assumption that there is something special about online terrorist-related speech such that it does not fit into existing frameworks by examining questions relating to the appropriate role, if any, of digital intermediaries in regulating online terrorist-related content and the extent to which proponents of human rights should be concerned with the free speech implications of intermediary liability. These issues are explored through a comparative examination of recent developments in Europe and the United States, including the shift from a voluntary to compulsory regulatory framework in the EU. The observations regarding the dangers to free speech that flow from aggressive efforts to regulate online terrorist-related expression are instructive for the broader question of the free speech implications of intermediary liability for extreme speech, which warrant attention in efforts to regulate 'hate speech' and disinformation on online platforms.

Chapter 5 tackles the challenges and dangers that flow from online disinformation from state actors. The primary objective of this chapter is to examine whether this type of extreme speech

has resulted in the failure of European and American free speech frameworks to adequately protect the right of citizens to unmanipulated factual information in the digital age. The harm with which this chapter is concerned is the damage to democratic norms and institutions that results from domestic disinformation campaigns from state actors. It argues that disinformation from state actors is the most dangerous form extreme speech in the digital age because it destabilises the liberal democratic order and, in so doing, hampers the ability of states to adequately protect vulnerable and historically marginalised groups from societal hatred and to safeguard the broader public from acts of terrorist violence.[71] It also foments the type of societal hatred and violence that regulatory efforts directed to 'hate speech' and terrorist-related expression are aimed at ameliorating.

The objective of **Chapter 6**, the final chapter, is to offer some observations and recommendations that emerge from the work undertaken in Part I and the preceding chapters in Part II. This chapter proposes, among other things, that it is time to re-examine the philosophical underpinnings of the American approach to free speech following the presidency of Donald Trump and the democratic backsliding that resulted from the online disinformation campaign targeting the integrity of the 2020 presidential election. It further argues that certain doctrines, including the USSC's government speech doctrine and broadly articulated immunity doctrines common in the United States and Europe, are not fit for purpose in the digital age. Human rights penality and the lessons to be learned from efforts to criminalise extreme speech in Europe is also discussed. Ultimately, this chapter concludes that while both approaches face challenges and warrant re-examining in light of the challenges brought by the digital age, the American approach is in most urgent need of reform given the ways in which domestic disinformation from state actors has functioned to swiftly erode democratic norms and institutions, while the lack of built-in protections against democratic backsliding that are integrated into the European framework have stymied proactive measures in the United States to acknowledge and address the danger posed by this type of expression.

The **Conclusion** to this thesis offers some final reflections that emerge from this comparative inquiry. These include the inherent challenges that arise from attempts to define types of

[71] See, e.g., Asif Efrat and others, 'Report on the Relationship Between Terrorist Threats and Governance Condition in the European Union' (*RECONNECT,* 29 June 2021) <https://reconnect-europe.eu/wp-content/uploads/2021/06/D11.3.pdf> accessed 1 July 2020, 15: 'Processes of democratic backsliding or outright "autocratization" bring countries into the "danger zone" of substantially increased probability of relatively high numbers of terrorist attacks and casualties.'

extreme speech that lie outside frameworks that provide expansive protection to freedom of expression, how this lack of clarity provides states with too much space to restrict core political speech, the ways in which power operates in this space, including decisions regarding who gets to speak in public discourse and who may legitimately promote violence and for what ends, and how to address the harm that results when the state itself is the purveyor of extreme speech. Areas for future research are also considered.

Finally, with all of the foregoing in mind, let us consider Ernest Young's observation that '[c]omparative law has long encouraged a healthy scepticism about deriving "answers" to one country's problems from another's experience', and that such scepticism 'counsels humility, not inattention'.[72] Thus, it is with both humility and attention that this inquiry begins.

[72] Ernest A Young, 'What Can Europe Tell Us About the Future of American Federalism?' (2017) 49 Ariz St L J 1109, 1110.

PART I

Introduction to Part I

While the objective of this thesis is evaluative in nature, that is, to examine the regulation of particular types of extreme speech in Europe and United States in order to explore broader questions concerning the appropriate limits on freedom of expression in liberal democracies in the digital age, the full benefit of a comparative analysis may only be achieved if supported by a sound doctrinal and philosophical foundation. Indeed, how we approach debates over free speech is largely influenced by the ways in which this right is conceptualised in a particular place.

The text of human rights instruments provide part of the overall picture of what free speech means in a given system. The interpretation of those texts by courts is also important, not only because it tells us what 'speech' is in a particular framework, but because it enriches our understanding of the connection between how this right is conceptualised in a given jurisdiction and the philosophical foundations of the right itself.[1] Additionally, philosophical arguments are relevant to constitutional interpretation because they assist courts in contextualising the right to freedom of expression within a particular legal framework.[2] For this reason, it is important to consider the extent to which broader human rights principles underpin specific rights within a particular system, and the ways in which courts are influenced by the broader philosophical principles that underpin the systems in which they are situated.[3]

Chapter 1 provides the doctrinal foundation for this thesis by examining the threshold questions of what constitutes speech within American and European constitutional frameworks, what classes of speech fall outside of such protections, and who is obligated to respect free speech rights in Europe and the United States. Chapter 2 builds on this foundation by examining the most salient ideological divergences in the European and American approaches to the regulation of extreme speech and the ways in which legal philosophy enriches our

[1] See Dieter Grimm, 'Free Speech in a Globalised World' in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy* (OUP 2010).
[2] Eric Barendt, *Freedom of Speech* (OUP 2005) 3 - 4.
[3] ibid.

understanding of why these divergences exist as well as the approaches themselves. Together, these chapters provide the doctrinal and philosophical grounding for this thesis.

Chapter 1

What is Speech? Delineating the Contours, Scope, and Limits of the Right to Freedom of Expression in Europe and the United States

'But in as much as the real justification of a rule of law, if there be one, is that it helps bring about a social end which we desire, it is no less necessary that those who make and develop the law should have those ends articulately in their minds.[1]

Introduction

While freedom of expression is commonly recognised as a fundamental right, entrenched in human rights instruments at the national and supranational levels, it is often articulated in general terms. As a result, it is largely left to courts to determine the contours and scope of free speech rights. Because the answers to difficult free speech questions are rarely provided by the text of human rights instruments, courts must decide them on the basis of the philosophical arguments that underlie the commitment to freedom of speech in a particular jurisdiction.[2] As discussed in the Introduction, while freedom of expression is an essential part of both American and European legal frameworks, the doctrinal foundations of free speech in the United States and Europe differ in meaningful ways that are often overlooked in the broader free speech discourse. Additionally, the discourse is often framed with abstract applications of a conception of rights that does not fully reflect their reach within a particular constitutional framework.

The objective of this chapter is to provide the doctrinal foundation for the comparative work undertaken in Part II by examining the threshold questions of what constitutes speech within

[1] Oliver Wendell Holmes, 'Law in Science and Science in Law' (1899) 12(7) Harv L Rev 443, 460.
[2] Eric Barendt, *Freedom of Speech* (OUP 2005) 7.

American and European constitutional frameworks, what classes of speech fall outside of such protections, and who is obligated to respect free speech rights. This exercise is guided by Mark Tushnet's advice to be particularly cognisant of the differences in the doctrinal structures of different constitutional frameworks to ensure that a comparative inquiry appreciates the nuances in each system.[3] While this examination is grounded in an analysis of the relevant jurisprudence of American and European courts – primarily the USSC and the ECtHR[4] - this is not a purely descriptive exercise as a doctrinal analysis reveals a degree of tension and contestation within judiciaries that reflect the broader challenges of balancing freedom and regulation in liberal democracies, particularly the challenges posed by extreme speech in the digital age.

This chapter begins by sketching the primary doctrinal principles upon which the USSC and ECtHR rely in determining the type of expression that falls within constitutional frameworks, including to what extent free speech principles apply to non-traditional forms of expression.[5] This is followed by an analysis of doctrinal exceptions to these principles, that is, the doctrinal basis for excluding expression that would otherwise warrant protection within constitutional frameworks. This analysis focuses on extreme speech and the extent to which both approaches are afflicted by methodological inconsistencies regarding the types of expression that fall outside of free speech protections and on what basis. Next, the question of who is obligated to respect free speech rights in American and European constitutional frameworks is considered. Finally, some observations concerning the ways in which the prior discussions serve to enhance our understanding of each approach as well as the broader questions that inform this thesis are provided.

I. What is 'Speech' and How is it Protected in American and European Constitutional Frameworks?

The USSC opines that the right to speak freely is regarded as 'the matrix, the indispensable condition of nearly every other form of freedom'.[6] Similarly, the ECtHR holds that freedom

[3] See Introduction, p 13.
[4] The judgments of lower federal courts in the United States and national courts in Europe are also considered when helpful in illustrating particular doctrinal principles.
[5] By 'constitutional framework' with respect to the ECtHR, I am referring to the case law of the Court, rather than the ways in which the ECHR may form part of Member States' national constitutional frameworks through incorporation into domestic law. See discussion of terminology in Introduction, pp 4 - 6.
[6] *Texas v Johnson*, 491 US 397, 414 (1989).

of expression is 'one of the essential foundations of [a democratic] society, one of the basic conditions for its progress and for the development of every man.'[7] While such broad pronouncements regarding the value of freedom of speech are useful in terms of discerning the general value afforded to this right within broader constitutional frameworks, they do not provide meaningful insights into what constitutes 'speech' within such frameworks, the ways in which courts distinguish between different classes of speech, and the relevance of such distinctions. Understanding these nuanced and often complex distinctions requires a more thorough analysis of the relevant jurisprudence, which is undertaken below.

A. How Courts Determine What Falls Within the Ambit of Free Speech Protections

The case law examined in this section primarily concerns the regulatory challenges posed by extreme speech. In ruling on challenges to interferences with expression, both the USSC and the ECtHR first consider whether relevant protections are triggered.[8] Following this initial determination, the approaches diverge; the USSC looks to the nature of the restriction to identify the appropriate test to apply for determining whether an interference is constitutionally permissible. Thus, the American approach to evaluating compliance with free speech protections hinges on the nature of a contested restriction.[9] The USSC's interpretation of the First Amendment generally prohibits the government from proscribing speech because of disapproval of the subject matter or viewpoint expressed. As a result, the government may generally not supress an entire category of speech, even if the regulation is viewpoint-neutral within that category.[10] In other words, the government may not distinguish favoured speech from disfavoured speech on the basis of the ideas or views expressed.[11] Accordingly, content-based restrictions on expression, including those restricting speech based on viewpoint, are presumptively unconstitutional and subjected to strict scrutiny, the USCC's most rigorous standard of review.[12] These doctrinal principles are underpinned by theories of democratic legitimacy and autonomy, which are explored in Chapter 2.

[7] *Handyside v United Kingdom* (1976) ECHR 5, para 49.
[8] The protections are the First Amendment and Article 10 of the ECHR, respectively.
[9] See e.g., *RAV v City of St. Paul*, 505 US 377, 388 (1992).
[10] See *Consolidated Edison Co v Public Serv Comm'n*, 447 US 530, 537 - 538 (1980).
[11] *Turner Broadcasting Sys, Inc v FCC*, 512 US 622, 643 (1994).
[12] *RAV* (n 9) 395. Strict scrutiny requires that the government demonstrate that a particular act is justified by a compelling interest and is narrowly drawn to serve that interest.

Content-neutral regulations are justified without reference to the content of the regulated speech.[13] The principal inquiry in determining whether a restriction on free speech is 'content-neutral' is whether the government has adopted the restriction because of its disagreement with the message that speech conveys.[14] A government regulation is 'content-neutral,' even though it has an incidental effect on some speakers or messages but not others, so long as it serves some purpose unrelated to the content of regulated speech.[15] This includes content-neutral restrictions that impose an incidental burden on speech.[16] A content-neutral regulation is subject to an intermediate (rather than strict) level of scrutiny, under which it will be upheld 'if it advances important governmental interests unrelated to the suppression of free speech and does not burden substantially more speech than necessary to further those interests'.[17] For example, it is a well-established principle that the government may impose reasonable time, place, and manner restrictions on expression so long as such restrictions are content-neutral.[18] Such restrictions include regulations requiring performers at concerts to use sound amplification equipment provided by a public authority in order to reduce excessive noise to the surrounding area.[19] This less rigorous form of scrutiny reflects a belief that content-neutral regulations pose a less substantial risk of excising unpopular ideas or viewpoints from public discourse.[20]

In stark contrast to the approach of the USSC, the ECtHR applies the same test to challenges to interferences with speech regardless of the nature of the impugned regulation. Moreover, there is no presumption against content and viewpoint-based restrictions. At this juncture, it is worth reiterating the institutional position of the ECtHR as compared to the USSC. As discussed in the Introduction, unlike the USSC, the judgments of which are directly effective and enforceable against individual states, the ECtHR requires the cooperation of Member States to secure the execution of its judgments.[21] In adjudicating on Article 10 cases, the ECtHR applies a balancing test that weighs the competing interests of the speaker with the stated aims of the government.[22] Specifically, the Court makes a determination as to whether

[13] *Clark v Community for Creative Non–Violence*, 468 US 288, 293 (1984).
[14] *Ward v Rock Against Racism*, 491 US 781, 791 (1989).
[15] *United States v O'Brien*, 391 US 367, 377 (1968).
[16] ibid.
[17] *Turner Broadcasting* (n 11).
[18] See *Ward* (n 14).
[19] ibid.
[20] *Turner Broadcasting* (n 11).
[21] See Introduction, pp 11 - 12.
[22] See, e.g., *Wingrove v United Kingdom*, App no 17419/90 (ECtHR, 25 November 1996).

an interference with an individual's Article 10 rights was necessary in a democratic society for achievement of one of the 'legitimate aims' identified in paragraph 2.[23] This involves a determination of whether the interference was prescribed by law,[24] pursues one of Paragraph 2's 'legitimate aims',[25] and was 'necessary in a democratic society' for achievement of said aims, that is, whether the interference was proportionate to the legitimate aim pursued and corresponded to a pressing social need.[26] Thus, an interference with an individual's right to freedom of expression only violates Article 10 if it fails to satisfy the requirements of Paragraph 2, that is, if the government's identified 'legitimate aim' does not prevail over the free speech interests of the speaker.

As touched on in the Introduction, in assessing whether such a 'need' exists in a given case, the ECtHR grants national authorities a 'margin of appreciation', which refers to the space that the ECtHR grants to national authorities in fulfilling their obligations under the Convention.[27] When regulating in the area of freedom of expression, a wide margin of appreciation is applied in matters of morals based on the ECtHR's conclusion that that there is no uniform conception of morals in Member States.[28] For this reason, a wider margin is applied in cases in which the restricted expression is deemed likely to 'offend intimate personal convictions within the sphere of morals or, especially, religion'.[29] With respect to speech concerning debates on public interest, the ECtHR applies a narrow margin of appreciation, which reflects the purported value the ECtHR places on political speech and the extent to which the scrutiny

---

[23] ibid.

[24] That is, whether the regulation was sufficiently clear and precise to enable a citizen to regulate his conduct in a way that is compatible with the law. See *Sunday Times v United Kingdom* App no 6538/74 (ECtHR, 26 April 1979).

[25] These include public safety, the protection of health or morals, and the protection of the reputation or rights of others.

[26] See, e.g., *Vejdeland v Sweden,* App no 1813/07 (ECtHR, 9 February 2012).

[27] See Introduction, pp 10 - 11. See also *Handyside* (n 7) para 49:

> By reason of their direct and continuous contact with the vital forces of their countries, State authorities are in principle in a better position than the international judge to give an opinion on the exact content of these requirements as well as on the 'necessity' of a 'restriction' or 'penalty' intended to meet them…

[28] *Gough v United Kingdom*, App no 49327/11 (ECtHR, 23 March 2015). See also *Müller and Others v Switzerland*, App no 10737/84 (ECtHR, 24 May 1988).

[29] *Wingrove* (n 22) para 58. See also *Ibragimov and Others v Russia*, App nos 1413/08 and 28621/11 (ECtHR, 4 February 2019), in which the court opined that:

> There is no uniform European conception of the requirements of "the protection of the rights of others" in relation to attacks on religious convictions. What is likely to cause substantial offence to persons of a particular religious persuasion will vary significantly from time to time and from place to place, especially in an era characterised by an ever-growing array of faiths and denominations. By reason of their direct and continuous contact with the vital forces of their countries, State authorities are in principle in a better position than the international judge to give an opinion on the exact content of these requirements with regard to the rights of others as well as on the "necessity" of a "restriction" intended to protect from such material those whose deepest feelings and convictions would be seriously offended.

applied by the Court varies depending on the type of speech at issue in a particular case. For example, it has held that 'there is little scope…for restrictions on political speech or debates on matters of public interest'.[30] The Court has also shown a willingness to grant a large measure of protection under Article 10 to media activities, opining in cases concerning the press, that the margin of appreciation is circumscribed by the interest of democratic society in ensuring and maintaining a free press.[31] Similarly, such an interest will weigh heavily in the balance in determining per paragraph 2, whether the restriction was proportionate to the legitimate aim pursued.[32]

The ECtHR's application of the margin of appreciation is the subject of widespread criticism from both scholars and judges for a lack of clarity and consistent application.[33] As acknowledged by Judge Lohmus in his dissent in *Wingrove, '*it is difficult to ascertain what principles determine the scope of th[e] margin of appreciation'.[34] Some scholars object to the doctrine on normative grounds. For example, George Letsas argues that the doctrine offends the values underlying human rights.[35] He contends that the 'doctrinal label of the margin of appreciation is shorthand for all the normative challenges that an international court, like Strasbourg, faces' and that 'we should move away from the unhelpful idea of discretion and the dangerous idea of deference'.[36] In the context of freedom of expression, some scholars argue that the margin of appreciation doctrine has functioned in practice in a manner that has resulted in an excessive degree of discretion to Member States in regulating extreme speech.[37] Other scholars argue that the Court does not adequately address the extent to which speech involving debates on public interest may offend the personal convictions or morals of the listeners.[38] The lack of engagement with the potential overlap between political speech and

[30] *Dichand and others v Austria*, App no 29271/95 (ECtHR, 26 May 2002) para 38.
[31] *Thoma v Luxembourg*, App no 38432/97 (ECtHR, 29 June 2001).
[32] ibid para 48.
[33] See, e.g., Ivan Hare, 'Extreme Speech Under International and Regional Human Rights Standards' in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy'* (OUP 2009); Janneke Gerards, 'Margin of Appreciation and Incrementalism in the Case Law of the European Court of Human Rights' (2018) 18 HRL Rev 495.
[34] *Wingrove* (n 22) para 1 (dissenting opinion of Judge Lohmus).
[35] George Letsas, 'The Margin of Appreciation Revisited' in Adam Etinson (ed), *Human Rights: Moral or Political* (OUP 2018). For a counterpoint to Letsas' argument, see Andreas Føllesdal,' Appreciating the Margin of Appreciation', in the same volume. See also Dominic McGoldrick, 'A Defence of the Margin of Appreciation and an Argument for its Application by the Human Rights Committee' (2016) 65(1) ICLQ 21.
[36] Ibid 309.
[37] See, e.g., Ian Cram, 'The Danish Cartoons, Offensive Expression, and Democratic Legitimacy' in Ivan Hare and James Weinstein (n 33).
[38] See, e.g., Jacob Mchangama and Natalie Alkiviadou, '"Hate Speech" Jurisprudence of the ECtHR through a Qualitative and Quantitative Lens' (*ECHR Blog,* 3 November 2020) <https://www.echrblog.com/2020/11/guest-post-hate-speech-jurisprudence-of.html> accessed 2 December 2020.

extreme speech is examined in Chapters 3 and 4 in relation to 'hate speech' and terrorist-related expression, respectively. Chapter 6 interrogates the contemporary application of the margin of appreciation doctrine in 'hate speech' cases.

B. Conduct as Speech: Defining the Contours and Limits of Free Speech Protections

Some of the most meaningful ways humans communicate are not through written or spoken words but, rather, through conduct. Indeed, an act can be so imbued with meaning that it communicates a belief or opinion far more powerfully than written or spoken expression. For example, in 2016 Colin Kaepernick ignited a national debate on racism and police brutality in the United States by kneeling during the national anthem before National Football League games.[39] Burning a national flag is banned in some European countries due to the message conveyed through the simple act of setting a flag on fire.[40] Accordingly, in order to fully appreciate the protection afforded to freedom expression in a particular place, the extent to which expressive conduct is protected warrants careful consideration.

In both the United States and Europe, free speech protections extend beyond written or spoken words to other mediums of expression. Both the USSC and the ECtHR subject challenges to regulations on conduct to a two-part analysis: first, they consider whether the conduct is sufficiently expressive to come within the ambit of free speech protections; second, they determine whether a challenged regulation violates such protections. This is where the similarities end, however, as the USSC and the ECtHR apply different tests to both parts of the analysis. The USSC has long held that the First Amendment protects certain forms of conduct. However, in so doing it recognises the First Amendment as endowing the government with a 'freer hand' in restricting expressive conduct than in restricting written and verbal expression.[41] In so doing, the Court carefully circumscribes protection for conduct and has expressly rejected the view that 'an apparently limitless variety of conduct can be labelled "speech" whenever the

[39] Billy Witz, 'This Time, Colin Kaepernick Takes a Stand by Kneeling,' *The New York Times* (1 September, 2016) <https://www.nytimes.com/2016/09/02/sports/football/colin-kaepernick-kneels-national-anthem-protest.html> accessed 12 July 2020.
[40] For example, the German Criminal Code proscribes destroying or damaging the German flag that is on public display (Strafgesetzbuch, Sec 90a). The penalty for this offense is imprisonment for a term exceeding five years or a fine. Flag burning is protected speech in the United States. The USSC has observed that '[t]he use of an emblem or flag to symbolize some system, idea, institution, or personality, is a short cut from mind to mind. Causes and nations, political parties, lodges and ecclesiastical groups seek to knit the loyalty of their followings to a flag or banner, a color or design' (*Johnson* (n 6) 405 (internal quotations and citations omitted)).
[41] ibid 6.

person engaging in the conduct intends thereby to express an idea'.[42] Instead, conduct qualifies as speech only when it is sufficiently imbued with elements of communication, that is, so long as it is 'inherently' expressive.[43] If, and only if, conduct is deemed inherently expressive does the First Amendment limit the authority of the government to restrict or compel it.[44]

In determining whether conduct is inherently expressive, the USSC considers whether the conduct was intended to be communicative and whether in context, it would reasonably be understood by viewers to be communicative.[45] The context includes the 'factual context and environment in which [the conduct] was undertaken'.[46] The presence of additional explanatory speech is 'strong evidence' that conduct is not so inherently expressive that it warrants First Amendment protection.[47] Applying this principle, the Court has recognised a wide array of conduct as expressive for First Amendment purposes, including burning the American flag,[48] politically motivated boycotts,[49] symbolic forms of protest,[50] and creating custom wedding cakes.[51]

Government regulations of inherently expressive conduct are subjected to a level of intermediate scrutiny referred to as the *O'Brien* test.[52] A regulation is sufficiently justified, i.e., constitutional, under the *O'Brien* test if it satisfies each of the following requirements: it is within the constitutional power of the government; it furthers an important or substantial governmental interest; the governmental interest is unrelated to the suppression of free expression; and the restriction on alleged First Amendment freedoms is no greater than essential to further the government's interest.[53]

---

[42] See *O'Brien* (n 15) 376.
[43] *Rumsfeld v Forum for Acad and Institutional Rights, Inc*, 547 US 47, 49 (2006).
[44] *Masterpiece Cakeshop, Ltd v Colorado Civil Rights Comm'n*, 138 SC 1719, 1742 (2018).
[45] *Johnson* (n 6).
[46] *Spence v Washington*, 418 US 405, 410 (1974) (explaining that the timing of conduct, during or around 'issues of great public moment,' may transform 'otherwise bizarre behavior' into conduct that 'the great majority of citizens' would understand 'the drift of'.). ibid.
[47] *Rumsfeld* (n 43) 66.
[48] See *Johnson* (n 6).
[49] See *NAACP v Claiborne Hardware Co*, 458 US 886 (1982).
[50] See *Tinker v Des Moines*, 393 US 503 (1969).
[51] *Masterpiece* (n 44).
[52] See *O'Brien* (n 15).
[53] ibid 377. This is effectively the same standard that courts apply to content-neutral regulations on speech such as time, place, and manner restrictions. See *Barnes v Glen Theatre, Inc*, 501 US 560, 566 (1991).

For those arguing that an interference with expressive conduct violates Article 10 of the ECHR, the approach of the ECtHR is both less and more onerous than the approach adopted by USSC. It is less onerous because it is easier to overcome the threshold determination of whether the conduct at issue falls within the ambit of Article 10. It is more onerous because once that threshold is met the test applied by the ECtHR is, overall, less protective of speech. Generally, the ECtHR holds that Article 10 extends beyond the substance of the ideas and information expressed to the form in which they are conveyed.[54] The term 'expression' has been widely construed to cover words, pictures, videos and conduct intended to convey an idea or information.[55] In deciding whether a certain act or conduct falls within the ambit of Article 10, the ECtHR assesses the nature of the act or conduct in question, in particular its expressive character seen from an objective point of view, as well as of the purpose or the intention of the person performing the act or carrying out the conduct in question.[56]

This assessment is relevant both to determining the applicability of Article 10 and assessing whether a violation has occurred in a given case. Using this framework, the ECtHR has recognised symbols and protest as falling within Article 10.[57] Unlike the USSC, the ECtHR applies the same balancing test in all Article 10 cases, regardless of whether the regulation at issue targets expressive conduct or more traditional forms of expression. That is, the court determines whether a regulation is prescribed by law, pursues a legitimate aim, and whether it is necessary in a democratic society for achievement of said aim, which includes the test of proportionality.

To appreciate the divergences in the protection afforded to expressive conduct in the United States versus Europe, an examination of the jurisprudence concerning free speech challenges to regulations on public nudity is instructive because the case law is extensive and offers insight into judicial reasoning in adjudicating free speech challenges to restrictions in this area. The USSC's jurisprudence in this area primarily addresses challenges to public nudity ordinances brought by establishments featuring nude dancing as a form of entertainment.[58] In these cases, the Court has etched out a set of principles that apply to First Amendment challenges to

[54] *Jersild v Denmark,* App no 15890/89 (ECtHR, 23 September 1994).
[55] *Murat Vural v Turkey*, App no 9540/07 (ECtHR, 21 January 2015).
[56] ibid.
[57] See *Donaldson v United Kingdom*, App no 56975/09 (ECtHR, 25 January 2011) and *Steel and Others v United Kingdom*, App nos 24838/94 (ECtHR, 23 September 1998), respectively.
[58] See, e.g., *California v LaRue*, 409 US 109 (1972) (holding that this form of entertainment might be entitled to First and Fourteenth Amendment protection under some circumstances.).

restrictions on public nudity more broadly and outside of the commercial context. First, while the Court holds that nudity is not an 'inherently expressive condition', it allows that it may, in some circumstances, constitute expressive conduct protected under the First Amendment.[59] However, it does not elaborate concerning what such circumstances might be. Second, it holds that government restrictions on public nudity, like other forms of expressive conduct, are evaluated using intermediate scrutiny.[60]

In adjudicating on challenges to government proscriptions on public nudity outside of the nude dancing context, lower courts apply the USSC's jurisprudence in this area. Accordingly, the first hurdle plaintiffs must overcome in such cases is to persuade the court that the conduct at issue is sufficiently expressive to trigger First Amendment protections. If the court finds in the plaintiff's favour on this point, it then applies intermediate scrutiny to determine whether the challenged regulation violates the First Amendment. In general, lower courts are reluctant to find that public nudity, even when engaged in for political purposes, is sufficiently expressive to fall within the ambit of the First Amendment. For example, in a case before the Seventh Circuit Court of Appeals, a woman brought an unsuccessful challenge to a citation for violating a city ordinance prohibiting public nudity.[61] The court was unpersuaded by the plaintiff's argument that baring her breasts in public as part of a 'Go Topless Day' for the purpose of protesting laws that prevented woman from appearing bare-chested in public was sufficiently expressive to trigger First Amendment protection.[62]

While the court did not question the veracity of the plaintiff's contention that her conduct was intended to be communicative, it found that she failed to 'offer facts from which it might reasonably be inferred that onlookers would have readily understood that her nudity was a political protest against the public-indecency ordinance.'[63] In its analysis, the court distinguished the plaintiff's case from USSC precedent involving flag burning, holding that while the expressive and overtly political nature of burning the American flag is intentional and 'overwhelmingly apparent', it was not so apparent that a woman's act of baring her breasts

[59] *City of Erie v Pap's AM*, 529 US 277 (2000) (holding that nude dancing for entertainment purposes is expressive conduct within the outer perimeters of the First Amendment.).
[60] ibid (opining that banning all public nudity, regardless of whether that nudity was accompanied by expressive activity, was content-neutral regulation and thus subject to less stringent *O'Brien* standard for evaluating restrictions on symbolic speech.).
[61] *Tagami v Chicago*, 875 F3d 375 (7th Cir 2017).
[62] ibid 377.
[63] ibid 378.

in public expressed a political message.[64] As further evidence for its conclusion, the court noted the presence of additional speech from plaintiff - that she appeared topless in public 'while expressing [her] views that women, like men, should not be prohibited from appearing bare chested in public' - was 'strong evidence' that her conduct was not sufficiently expressive to warrant First Amendment protection.[65]

A case from the Ninth Circuit Court of Appeals offers an example of how courts apply intermediate scrutiny to First Amendment challenges to government proscriptions on public nudity.[66] The plaintiffs in this case were self-described 'body freedom advocates' who ran afoul of San Francisco's public nudity ordinance by appearing naked at political rallies and other city events. The court found that even if the plaintiffs' public nudity was sufficiently expressive to warrant First Amendment protection, the challenged ordinance was a valid, content-neutral regulation under the *O'Brien* test.[67] First, the court held that restricting public nudity fell within the city's traditional police powers. Second, it held that the ordinance furthered the city's important and substantial interest in protecting those who were unwillingly or unexpectedly exposed to public nudity and prevented incidents that interfered with the safety and free flow of foot and vehicle traffic.[68] Third, it held that the city's interest was unrelated to the suppression of free expression because the ordinance 'regulates public nudity whether or not it is expressive'.[69] That is, the ordinance was aimed 'at the conduct itself, rather than at the message conveyed by that conduct'.[70] Finally, the court held that the incidental restriction on alleged First Amendment freedoms was no greater than what was necessary to further the city's interests.[71]

---

[64] ibid.

[65] ibid. Judge Rovner's dissent challenges the majority's conclusion that plaintiff Tagami's act of baring her breasts was not sufficiently communicative:

> Tagami was not sunbathing topless to even her tan lines, swinging topless on a light post to earn money, streaking across a football field to appear on television, or even nursing a baby (conduct that is exempted from the reach of the ordinance). Her conduct had but one purpose - to engage in a political protest challenging the City's ordinance on indecent exposure. Tagami engaged in the paradigm of First Amendment speech - a public protest on public land in which the participants sought to change a law that, on its face, treats women differently than men. It is difficult to imagine conduct more directly linked to the message than that in which Tagami engaged. ibid 381.

[66] *Taub v San Francisco*, 696 FedAppx 181 (9th Cir 2017).

[67] ibid 182.

[68] ibid 183.

[69] ibid.

[70] ibid.

[71] ibid.

While the ECtHR's jurisprudence concerning free speech challenges to restrictions on public nudity is far less extensive than the USSC, it is instructive in elucidating the distinctions in the American and European approaches to the regulation of expressive conduct. In the case of *Gough v United Kingdom*, the ECtHR considered the arrest, prosecution, conviction, and imprisonment of a man for numerous incidents of public nudity.[72] The Court accepted the applicant's declaration that his public nudity gave expression to his opinion as to the 'inoffensive nature of the human body' that, in turn, gave rise to a belief in social nudity that he expressed by appearing naked in public. Based on these facts, the Court was satisfied that the applicant's public nudity constituted a form of expression that fell within the ambit of Article 10 and that the government's actions constituted an interference with that form of expression. Unlike the USSC, the ECtHR did not place any burden on the applicant to demonstrate that those viewing him in a state of nudity would understand the idea he intended to express in order to trigger free speech protections.

In its Article 10 analysis, the ECtHR disposed of the first two elements efficiently, holding that the public order offence was sufficiently precise to provide reasonable foreseeability that public nudity would fall within its remit and that its aim to prevent disorder and crime was a legitimate aim under paragraph 2. In considering the final element, the Court applied a wide margin of appreciation in finding that the interference met a pressing social need in response to the repeated 'anti-social' conduct of the applicant. It further held that the measures taken by authorities were proportionate to the aim of the prevention of disorder and crime and that Article 10 did not go so far as to enable individuals 'sincerely convinced of the virtue of their own beliefs' to impose antisocial conduct on unwilling members of society'.[73]

What broader observations regarding the protection afforded to expressive conduct may be gleaned from a comparative analysis of jurisprudence concerning proscriptions on public nudity? First the threshold to satisfy whether conduct is sufficiently expressive to fall within the ambit of free speech protections is significantly lower for the ECtHR. Effectively, all an applicant must show is a genuine belief that the conduct represents the expression of a particular idea or belief. There is no additional burden on the applicant to establish that onlookers would readily understand the expressive nature of the conduct at issue. In this way,

[72] *Gough* (n 28). The offence at issue in this case was a breach of the peace.
[73] ibid para 184.

the ECtHR's approach may be regarded as more protective of expression than the USSC with respect to the first stage of the analysis, that is, whether conduct falls within the ambit of free speech protections. However, with respect to whether an interference with sufficiently expressive conduct is permissible, the ECtHR's approach is less protective of speech than the USSC in that the scrutiny undertaken is less rigorous. For example, there is no consideration of whether the challenged regulation specifically targets the content of the speech or the viewpoint expressed, and there is no onus placed on the government to enact regulations that are content and viewpoint neutral.[74] Instead, all regulations are reviewed using the same level of scrutiny. The protection afforded to expressive conduct is relevant to the examination of 'hate speech' in Chapter 3.

II. What Classes of Expression are Excluded from American and European Free Speech Frameworks?

A comprehensive understanding of the right to freedom of expression in a given jurisdiction requires an appreciation not only of the type of speech that falls within free speech frameworks, but also of the speech that courts determine lie outside of such protections, i.e., doctrinal exceptions to general free speech principles. This is particularly true in jurisdictions that emphasise the importance of freedom of expression, in which exceptions to free speech protections in the realm of public discourse require particularised and persuasive justifications. An analysis of the exceptions to the broad protections afforded to freedom of expression in the USSC and the ECtHR highlight the significant challenges in developing doctrinal frameworks that are, on the one hand, rooted in a firm commitment to freedom of expression and, on the other hand, carve out exceptions for certain forms of objectionable speech. Such analysis also exposes intra-jurisdictional normative debates among USSC justices regarding the proper scope of protection for extreme speech, in particular, revealing a degree of contestation and discord within the judiciary that is often overlooked in examinations of American free speech doctrine.

A. Exceptions to Free Speech Protections in the United States

[74] ibid.

Over the course of several iterations, the USSC has carved out exceptions to First Amendment protections for particular categories of speech.[75] These include, inter alia, incitement to violence[76], true threats,[77] and fighting words[78]. An analysis of the relevant jurisprudence exposes a doctrinal quagmire that reflects the inherent challenges in delineating exceptions within a framework premised on expansive protection of speech in public discourse irrespective of content or viewpoint. Inconsistencies in the Court's jurisprudence, while largely glossed over in majority opinions, are interrogated in impassioned concurrences and dissents, which highlight important debates concerning the extent to which, and on what basis, the government may regulate certain forms of extreme speech.

An examination of the USSC's jurisprudence involving fighting words provides a useful illustration of these debates. In *R.A.V. v. City of Saint Paul*, the petitioner, along with several others, allegedly burned a cross in the yard of an African American family. While this conduct violated a number of laws, the city of Saint Paul chose to charge the petitioner with violating the city's 'Bias-Motivated' Crime Ordinance, which prohibited the display of a symbol that aroused 'anger, alarm or resentment in others on the basis of race, color, creed, religion or gender'.[79] The majority opinion hinged on the fact that the Ordinance, though addressing a categorical exclusion to the general rule against content-based restrictions in the form of fighting words, went beyond 'mere content discrimination' to viewpoint discrimination in permitting displays containing fighting words that did not invoke the proscribed characteristics.[80]

[75] The USSC describes these exceptions as 'well-defined and narrowly limited classes of speech'. See *Chaplinsky v New Hampshire*, 315 US 568, 571 - 572 (1942). Recently, the Court refused the government's appeal to add depictions of animal cruelty to this list. *United States v Stevens*, 130 SCt 1577 (2010).

[76] See *Brandenburg v Ohio*, 395 US 444, 447 (1969) (holding that in order for the government to proscribe expression advocating for lawless action, it must establish that particular speech is both 'directed to inciting or producing imminent lawless action' *and* is 'likely to incite or produce such action'.). Issues concerning the relationship between incitement and proscriptions on extreme speech are examined in Chapters 3 and 4 in relation to 'hate speech' and terrorist-related expression, respectively.

[77] See *Virginia v Black*, 538 US 343, 359 (2003) (holding that '"[t]rue threats" encompass those statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals'.).

[78] See *RAV* (n 9). See also *Cohen v California,* 403 US 15, 20 (1971) (holding that fighting words are 'those personally abusive epithets which, when addressed to the ordinary citizen, are, as a matter of common knowledge, inherently likely to provoke violent reaction'.).

[79] ibid (*RAV*) 380.

[80] For example, the Court highlighted that fighting words in connection with other ideas, such as to express hostility to homosexuality, were not covered by the ordinance. ibid 391.

The majority held that the Court's precedent expressly stating that 'the protection of the First Amendment does not extend' to particular categories of speech, including fighting words, *actually* meant that these categories were 'not entirely invisible to the Constitution'.[81] In so doing, it delineated an exception to the rule that content-based discrimination was presumptively unconstitutional involving circumstances '[w]hen the basis for the content discrimination consists entirely of the very reason the entire class of speech...is proscribable.'[82] In applying this rule to the Ordinance, the Court reasoned as follows: fighting words are categorically excluded from First Amendment protections because their content embodies a 'particularly intolerable (and socially unnecessary) *mode* of expressing *whatever* idea the speaker wishes and convey' rather than that their content communicates a particular idea.[83] Thus, while the government may choose to single out fighting words as an especially offensive *mode* of expression, it may not proscribe a subclass of fighting words that express certain messages of intolerance, as this reflects an effort to 'handicap the expression of particular ideas'.[84] The Court rejected the city's argument that only a content-based measure would communicate to minority groups that the 'group hatred' aspect of such speech was 'not condoned by the majority', opining that '[t]he point of the First Amendment is that majority preferences must be expressed in some fashion other than silencing speech on the basis of its content'.[85]

The vigorous and, at times, caustic concurring opinions in *R.A.V.* highlight the doctrinal revisionism engaged in by the majority as well as the methodological inconsistencies in its discussion of the nature of, and basis for, categorical exceptions to First Amendment protections.[86] For example, Justice White accuses the majority of ignoring the Court's long-standing precedent permitting content-based restrictions, without qualification, on speech in 'a few limited areas that are of such slight social value as a step to any truth that any benefit derives from them is clearly outweighed by the social interest in order and morality.'[87] He further admonishes the majority's renunciation of the long standing principle that categorical exceptions are not within the area of constitutionally protected speech as 'not literally true',

[81] ibid 383.
[82] ibid 388.
[83] ibid 393.
[84] ibid.
[85] ibid 392. Justices Blackmun and O'Connor joined White's concurrence (along with Justice Stevens, except as to part I-A).
[86] Justice White begins his concurrence by stating 'I agree with the majority that the judgment of the Minnesota Supreme Court should be reversed. However, our agreement ends there'. ibid 39.
[87] ibid 399.

arguing that the majority erred in holding that the First Amendment extends to categories of expression long considered *undeserving* of such protection. He argues that it is inconsistent for the Court to hold, on the one hand, that the government may proscribe an entire category of expression because the content is evil but may not, on the other hand, treat a subset of that category differently based on that same principle.[88] He also expresses particular concern with the majority's suggestion that expressions of violence in the form of intimidation and racial hatred are 'of sufficient value to outweigh the social interest in order and morality that has traditionally placed such fighting words outside the First Amendment', and argues that by characterising fighting words as a form of debate, the majority 'legitimates hate speech as a form of public discussion'.[89] He concludes by stating that a ban on a subset of fighting words would be appropriate as restricting 'only the social evil of hate speech' without creating a danger of driving viewpoints from the marketplace.[90]

In a separate and equally acerbic concurrence, Justice Stevens characterises the majority's opinion as 'an adventure in a doctrinal wonderland' that 'disrupts well-settled principles of First Amendment law'.[91] While Justice Stevens agrees with much of Justice White's analysis, he wrote separately to express his concerns regarding the Court's 'categorical approach' to the First Amendment.[92] Like Justice White, he takes issue with the majority's revision of the long standing categorical approach to speech regulation. However, unlike Justice White, he believes that the categorical approach itself is 'is unworkable and the quest for absolute categories of "protected" and "unprotected" speech ultimately futile'.[93] Referencing precedent relating to fighting words and child pornography, Justice Stevens emphasises that whether speech falls into a category of unprotected speech is righty determined in part by its content and, as a result, the majority erred in suggesting that regulation cannot be predicated thereon. He references several examples of the Court upholding content-based restrictions on speech, including zoning ordinances that regulate movie theatres based on the content of the films shown, a restriction on the broadcast of specific indecent words, a state law that restricted the speech of state

[88] ibid 401.
[89] ibid 402.
[90] He also concludes that the ordinance would survive strict scrutiny because it 'helps to ensure the basic human rights of members of groups that have historically been subjected to discrimination…'. ibid 395.
[91] Justices White and Blackmun joined as to Part I of Stevens' concurrence.
[92] ibid 417.
[93] ibid 427.

employees only as concerned partisan political matters, and the regulation of misleading advertising and labelling.[94]

Additionally, Justice Stevens takes issue with the Court's creation of a 'rough hierarchy' in the constitutional protection of speech, with core political speech at the top, followed by commercial speech and non-obscene, and sexually explicit speech as 'a sort of second-class' expression, and obscenity and fighting words at the bottom, receiving 'the least protection of all'.[95] He argues that in ruling that proscribable speech cannot be regulated based on subject matter, the Court elevates fighting words and obscenity to the same level of protection afforded to commercial speech, and perhaps even core political speech: '[I]f Congress can prohibit false advertising directed at airline passengers without also prohibiting false advertising directed at bus passengers, it is ironic to hold that a city cannot regulate fighting words based on 'race, color' creed, religion, or gender' while leaving unregulated fighting words' based on other characteristics.[96] Justice Stevens argues that, in so doing, the Court 'turns First Amendment law on its head: Communication that was once entirely unprotective…is not entitled to greater protection than commercial speech – and possibly greater protection than core political speech'.[97]

The concurring opinions of Justices White and Stevens highlight the extent to which the USSC's approach to the regulation of speech in public discourse and, in particular, to categorical exclusions in this area, is fraught with contradictions and contestation. These contradictions concern fundamental questions regarding the extent to which certain types of speech in public discourse are deserving of protection from government interference, as well as the inherent challenges in delineating exceptions to broadly articulated free speech principles. In so doing, the Justices correctly emphasise the ways in which the *R.A.V.* majority makes no effort whatsoever to reconcile conflicting precedent, such as that the United States 'has permitted the restrictions upon the content of speech in a few limited areas, which are of such slight social value as a step to the truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality', with the broader principles underpinning the First Amendment, including that the government may not regulate speech

[94] ibid 421.
[95] ibid 422.
[96] ibid.
[97] ibid 423.

based on the purported value it adds to public discourse and that speech is deserving of full First Amendment protection regardless of whether it survives an ad hoc balancing of relative social costs and benefits.[98] Additionally, the majority's emphasis on the social value of certain categories of speech belies the Court's long-standing principle that the state may not proscribe speech in public discourse on the basis that it is 'not worth it'.[99]

While the contemporary USSC continues to provide robust protections to free speech, especially when compared with other liberal democracies, these doctrinal contradictions and tensions endure, suggesting that the landscape of free speech in the United States is less settled than it may appear on cursory review. The ways in which online speech is further disrupting this landscape are examined in Part II.

B. Exceptions to Free Speech Protections in Europe

While the ECtHR, like the USSC, permits exceptions to free speech protections for particular classes of expression, the doctrinal basis for these exceptions highlights the extent to which the European and American approaches are grounded in distinct conceptions concerning the intrinsic value of speech in democratic societies. For example, while the ECtHR attaches strong value to the notion of public debate in cases challenging speech regulations, this is just one of many factors it considers and, unlike the USSC, it does not generally place a predominant emphasis on the social benefits deriving from open and unfettered speech in public discourse.

In another important divergence from American free speech doctrine, the ECtHR holds that not all speech is worthy of inclusion in public debate and the State may and should act as the arbiter in determinations of such worthiness. By way of example, the ECtHR holds that Member States may restrict speech that does 'not contribute to any form of public debate capable of furthering progress in human affairs.'[100] Similarly, the Council of Europe expressly eschews 'absolute liberalism' in the realm of speech regulation in favour of an approach predicated on the notion that not all ideas are deserving of circulation and that the right to

[98] See *Stevens* (n 75) 1585.
[99] ibid.
[100] *Otto-Preminger-Institut v Austria*, App no 13470/87 (ECtHR, 20 September 1994) para 49.

express one's ideas may be outweighed by competing societal interests.[101] Thus, while the ECtHR holds that an individual taking part in a public debate on a matter of general concern is permitted a degree of exaggeration or even provocation, that is, 'to make somewhat immoderate statements,' it imposes significant limitations on what types of ideas may be permitted in public discourse as well as the *ways* in which such ideas may be expressed by imposing duties and responsibilities on the exercise of Article 10 rights.[102] The philosophical underpinnings of imposing duties and responsibilities on the exercise of rights are examined in Chapter 2.

In delineating exceptions to free speech principles for particular classes of speech, the ECtHR looks to Article 17 of the Convention, which serves as the Convention's 'anti-abuse clause'.[103] The idea of an abuse clause in international human rights law originated shortly after World War II, and complemented efforts at the national level in Europe to address the concerns of those states that had confronted Nazism and fascism with combating the 'enemies of democracy'.[104] It was created to resist the revival of totalitarianism in Europe and to exclude protection for anti-democrat activity.[105] Thus, Article 17 may be regarded as a reflection of the political climate that governed Europe at the time of its adoption, and represents the response of European democracies to fascist and communist threats.[106]

Article 17 provides that the ECHR does not imply a right for any state, group or person 'to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the

[101] See 'Report on the Relationship between Freedom of Expression and Freedom of Religion: the Issue of Regulation and Prosecution of Blasphemy, Religious Insult and Incitement to Religious Hatred' (*Venice Commission*, 23 October 2008) <https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2008)026-e> accessed 10 July 2020.
[102] See, e.g., *Kuliś and Różycki v Poland*, App no 27209/03 (ECtHR, 6 January 2010) para 39; see also *Mamère v France,* App no 12697/03 (ECtHR, 17 October 2006) para 25.
[103] For a comprehensive overview of the ECtHR's Article 17 jurisprudence, see 'Guide on Article 17 of the European Convention on Human Rights: Protection of abuse of rights' (*European Court of Human Rights*, updated on 31 August 2019) available at <https://www.echr.coe.int/Documents/Guide_Art_17_ENG.pdf> accessed 10 July 2020.
[104] See Hannes Cannie and Dirk Voorhoof, 'The Abuse Clause and Freedom of Expression in the European Convention on Human Rights: An Added Value for Democracy and Human Rights Protection?' (2011) 29(1) NQHR 54.
[105] Mark E Villiger, 'Article 17: ECHR and Freedom of Speech in Strasbourg Practice' in Joseph Casadevall and others (eds), *Freedom of Expression: Essays in Honour of Nicolas Bratza* (Wolf Legal Publishers 2012) 321-322.
[106] ibid 57. See also Pauline de Morree, *Rights and Wrongs under the ECHR: The prohibition of abuse of rights in Article 17 of the European Convention on Human Rights* (Intersentia 2016) 23: '[Article 17] is the most explicit expression of the ambition of the Convention as a whole: preventing the emergence or re-emergence of totalitarian regimes in Western Europe'.

Convention'.[107] The ECtHR interprets Article 17's purpose as 'to make it impossible for [groups or individuals] to derive from the Convention a right to engage in any activity or perform any act aimed at destroying any of the rights and freedoms set forth in the Convention' and that 'no person may be able to take advantage of the provisions of the Convention to perform acts aimed at destroying the aforesaid rights and freedoms…'[108] This includes circumstances in which speech is directed against the Convention's underlying values, including justice and peace,[109] effective political democracy,[110] peaceful settlement of international conflicts and sanctity of human life,[111] tolerance, social peace, and non-discrimination[112], and gender equality.[113] In such cases, the ECtHR may declare an applicant's Article 10 complaint incompatible *ratione materiae* with the Convention.[114]

The ECtHR directs that Article 17 should only be used on an 'exceptional basis and in extreme cases'[115] if is 'immediately clear' that the speech at issue sought to 'deflect [Article 10] from its real purpose by employing the right to freedom of expression for ends clearly contrary to the values of the convention'.[116] The application of Article 17 in Article 10 cases extends beyond 'explicit and direct remarks that do not require any interpretation' and may include satirical and artistic modes of expression that the Court determines function as a 'guise' for dangerous speech.[117] The Court has interpreted Article 17 to include certain categories of speech, including incitement to violence,[118] support for terrorist activity,[119] incitement to hatred,[120] anti-Semitism and holocaust denial,[121] attacks against religious groups,[122] advocating for national socialism[123] and, more broadly, 'hate speech'.[124] In this way, Article 17 serves a

[107] ECHR art 17.
[108] *Lawless v Ireland (No. 3*), App no 332/57 (ECtHR, 1 July 1961) para 7.
[109] *Garaudy v France*, App no 65831/01 (ECtHR, 24 June 2003).
[110] *Refah Partisi (The Welfare Party) and others v Turkey*, App nos 41340/98, 41342/98, 41343/98 and 41344/98 (ECtHR, 13 February 2003).
[111] *Hizb ut-Tahrir and others v Germany*, App no 31098/08 (ECtHR, 12 June 2012).
[112] See *Pavel Ivanov v Russia*, App no 35222/04 (ECtHR, 20 February 2007); *Norwood v United Kingdom*, App no 23131/03 (ECtHR, 16 November 2004).
[113] See *Kasymakhunov and Saybatalov v Russia*, App nos 26261/05 and 26377/06 (ECtHR, 14 June 2013).
[114] See *M'Bala v France*, App no 25239/13 (ECtHR, 20 October 2015).
[115] *Paksas v Lithuania*, App no 34932/04 (ECtHR, 6 January 2011) para 87.
[116] *Ibragimov* (n 29) para 62.
[117] See *M'Bala* (n 114) para 40.
[118] *Hizb ut-Tahrir* (n 111).
[119] *ROJ TV A/S v Denmark,* App no 24683/13 (ECtHR, 17 April 2018).
[120] *Garaudy* (n 109).
[121] *M'Bala* (n 114).
[122] *Norwood* (n 112).
[123] *Kühnen v Federal Republic of Germany*, App no 12194/86 (ECtHR, 12 May 1988).
[124] See *Jersild* (n 54). The unique challenges posed by the regulation of 'hate speech' are discussed in Chapter 3

similar function to the categorical exclusions to First Amendment protection developed by the USSC in that both are built into doctrinal frameworks for the purpose of providing courts with a mechanism to remove certain types of expression from the ambit of free speech protections. While the USSC's list of excluded speech derives from its precedent concerning the nature and scope of the First Amendment, the ECtHR relies on its interpretation of Article 17 within the context of the underlying principles and objectives of the Convention.

The Court's application of Article 17 in Article 10 cases has faced legitimate criticism on the basis that excluding certain classes of speech from Article 10's prima facie protection is inconsistent with the Court's general approach of providing broad protection to freedom of expression.[125] While it is reasonable to assume that the practical effect of applying Article 17 rather than Article 10 in a given case would lead to the same outcome, that is, an interference with speech considered sufficiently extreme to fall within the ambit of Article 17 would also be considered proportionate under the Court's balancing test, the analysis should not end here. Indeed, the Court's application of Article 17 in extreme speech cases is fundamentally problematic because, among other things, it permits the ECtHR to sidestep constitutional protections for freedom of expression, thereby eliminating significant procedural safeguards for applicants requesting relief from interferences with their Article 10 rights.[126]

In addition to the lack of adequate procedural safeguards, the ECtHR's application of Article 17 in Article 10 cases is problematic because it has led to doctrinal uncertainty. Put simply, it is impossible to predict the factual circumstances in which the ECtHR will apply Article 17 in lieu of Article 10.[127] For example, notwithstanding the Court's pronouncement that holocaust denial falls into the 'category of prohibited aims under Article 17' on the basis of its incompatibility with democracy and human rights, in cases involving holocaust denial, decisions concerning whether to apply Article 10, in which Article 17 is sometimes but not always used an as aid in the interpretation of paragraph 2[128], or Article 17 to declare an application as incompatible *ratione materiae*,[129] are 'taken on a case-by-case basis and will depend on all the circumstances of each individual case'.[130] Additionally, while the ECtHR

[125] See e.g., Cannie (n 104); Hare (n 37) 78 - 79.
[126] ibid (Cannie).
[127] For a comprehensive examination of the inconsistencies in the ECtHR's approach to Article 17 in Article 10 cases, see Morree (n 106).
[128] See, e.g., *Williamson v Germany*, App no 64496/17 (ECtHR, 8 January 2019).
[129] See *Perinçek v Switzerland*, App no 27510/08 (ECtHR, 15 October 2015).
[130] *Pastörs v Germany*, App no 55225/14 (ECtHR, 3 October 2019) para 37.

identifies support for terrorism as a class of speech that triggers Article 17, it has failed to reference Article 17 in decisions involving such speech. For example, in a 2018 case, the Court did not make reference to Article 17 despite its determination that the impugned statements, which followed a terrorist attack, 'praised warlords as "heroes" and "patriots"' who showed 'an example of how one should fight against Russia' and 'glorified terrorism'.[131] Further adding to the confusion, in the 2020 case of *Atamanchuk v Russia*, the Court engaged in an Article 10 analysis *before* making a determination regarding the applicability of Article 17.[132] In his concurring opinion, Judge Lemmens highlighted the lack of coherency in the majority's approach, stating that:

> Following a strict logic, the Court would first have to…decide whether the complaint was admissible or not. If the Court found that the statements made by the applicant were covered by Article 17, then Article10 would have to be declared inapplicable and the complaint incompatible *ratione materiae* with the Convention, without there being any need to examine whether the interference with the applicant's freedom of expression was lawful, pursued a legitimate aim, and was proportionate to that aim. If, by contrast, the Court were to find that the applicant's statements were not such that they were covered by Article 17, then it would have to declare Article 10 applicable and (unless the complaint had to be declared inadmissible on another ground) proceed with an examination of the merits. In the present case, the Court in effect leaves open the question whether the Article 10 complaint is admissible.[133]

Ultimately, the Court's Article 17 jurisprudence in this area may reasonably be regarded as a patchwork of decisions that, when considered collectively, fail to provide a consistent and coherent approach to the application of Article 17 in cases involving challenges to restrictions on Article 10 rights, particularly in cases involving extreme speech. Thus, while the application of Article 17 is problematic given the broader constitutional framework in which it is situated, which considers speech in public discourse worthy of the highest level of protection, is also reflects the difficulties of delineating exceptions to broadly articulated free speech rights.

While the differences between the American and European approaches to speech regulation receive considerable attention in academic scholarship, an examination of efforts by American and European courts to craft doctrinal exceptions to free speech protections reveal similar challenges in developing consistent and methodologically coherent jurisprudence in this area.

---

[131] *Stomakhin v Russia*, App no 52273/07 (ECtHR, 8 October 2018) para 104.
[132] *Atamanchuk v Russia*, App no 4493/11 (ECtHR, 12 October 2020).
[133] ibid para 2.

This is particularly apparent in the context of the regulation of extreme speech. A detailed analysis of USSC jurisprudence reveals the degree of judicial discord concerning the extent to which particular categories of expression are deserving of protection and the normative basis on which to make such distinctions. Unlike the USSC, the ECtHR does not place particular emphasis on delineating categorical exclusions to Article 10 but, rather, balances the right to freedom of expression with the limitations of paragraph 2. To the extent it has undertaken attempts to identify certain categories of speech that lie outside of Article 10's protections, such efforts have led to methodological uncertainty as there is no clearly delineated test for determining the circumstances in which it is appropriate to apply Article 17 in lieu of Article 10. Thus, while the USSC struggles with developing a consistent approach to applying clearly delineated categorical exclusions to its First Amendment doctrine, the ECtHR has, to date, failed to develop a consistent or coherent approach for its use of Article 17 in free speech cases. Thus, both approaches struggle to apply exclusions to free speech protections in a coherent and consistent manner. The ECtHR's application of Article 17 is further examined in the context of 'hate speech' and terrorist-related expression in Chapters 3 and 4, respectively, and in Chapter 6.

III. Who is Obligated to Respect Free Speech Rights?

The final threshold free speech question this chapter considers is who is responsible for respecting free speech rights? Constitutional theory traditionally associates the application of fundamental rights with obligations on a state.[134] This means that actions to enforce fundamental rights in traditional constitutional frameworks may generally only be brought against state actors, with private actors having no obligations or responsibilities to uphold or promote the realisation of such rights.[135] The extent to which a particular jurisdiction conforms to these traditional conceptions directly impacts the extent to which the application of fundamental rights extends to private actors in relation to one another. This section examines the primary doctrinal principles in the United States and Europe that inform this question.

[134] See Eleni Frantziou, *The Horizontal Effect of Fundamental Rights in the European Union: A Constitutional Analysis* (OUP 2019).
[135] See Aoife Nolan, 'Holding non-state actors to account for constitutional economic and social rights violations: Experience and lessons from South Africa and Ireland' (2014) 12(1) ICON 61.

The principles discussed in this section, in addition to answering the broader question of who is responsible for respecting fundamental rights in American and European constitutional frameworks, inform other issues explored in this thesis, including the philosophical inquiries undertaken in Chapter 2. These include how differing conceptions of liberty impact the scope and application of free speech principles in American and European constitutional frameworks, whether it is appropriate to place positive obligations on the government to protect fundamental rights, and whether and to what extent citizens undertake responsibilities as a result of exercising rights.

A. Obligations to Respect Free Rights in the United States: State Action and Public Forum Doctrines

The American constitutional framework embraces traditional conceptions of constitutional theory that associate the application of fundamental rights with state obligations. This may be explained, in part, by the fact that the United States Constitution is a charter of purely negative liberties, that is, it tells the state to leave people alone.[136] This reflects the steadfast commitment to individualism and long-standing tradition of negative freedom that lie at the heart of American legal and cultural identity. For this reason, the rights guaranteed in the Bill of Rights - the first ten Amendments to the United States Constitution - only protect rights holders from the acts of government, not those of private actors.[137] Additionally, the Constitution does not impose upon the government an obligation to protect the life, liberty, or property of citizens from invasion by private actors.[138]

The USSC has crafted two distinct doctrines that reflect the ways in which these traditional conceptions of fundamental rights apply in cases alleging violations of First Amendment rights. The state action doctrine provides the answer to the question of whether challenged conduct

[136] See, e.g., *Harris v McRae*, 448 US 297, 318 (1980) (holding that'[a]lthough the liberty protected by the Due Process Clause affords protection against unwarranted government interference with freedom of choice in the context of certain personal decisions, it does not confer an entitlement to such funds as may be necessary to realise all the advantages of that freedom. To hold otherwise would mark a drastic change in our understanding of the Constitution'.); *Bowers v Devito* 686 F 2d 616, 618 (7th Cir 1982) (holding that 'because the bill of rights is a charter of negative liberties, the state need not protect people from danger'.). See also Justice Brennan's dissent in *Deshaney v Winnebago County Dep't of Soc. Services*, 489 US 189, 204 (1989) (opining that the USSC's 'baseline is the absence of positive rights in the Constitution and a concomitant suspicion of any claim that seems to depend on such rights'.).

[137] *Hudgeons v NLRB*, 424 US 507, 513 (1976). See also *Manhattan Cmty Access Corp v Halleck*, 139 SCt 1921, 1928 (2019) (holding that '[t]he Free Speech Clause prohibits only *governmental* abridgment of speech…The Free Speech Clause does not prohibit *private* abridgment of speech'.).

[138] *Castle Rock v Gonzales*, 545 US 748, 755 (2005).

may be attributed to state or private actors. The public forum doctrine instructs courts as to what level of protection from government interference contested speech attributed to the state warrants in a particular case.

1. The State Action Doctrine

The state action doctrine draws the line between governmental and private conduct.[139] Accordingly, in each case in which the First Amendment is implicated, a court must determine whether a challenged act is that of the government, i.e., whether a particular restriction is sufficiently governmental in character to constitute 'state action'.[140] The USSC opines that the state action doctrine protects 'a robust sphere of individual liberty' by distinguishing the government from individuals and private entities and by enforcing the boundary between the governmental and the private.[141]

While the USSC has devised several tests for determining whether there is sufficient state involvement for a finding of state action in a given case, it emphasises that the relevant issue is whether such power has been exercised rather than the form in which it has been applied,[142] and that the overall inquiry is whether there is an adequate nexus between the private behaviour and the state.[143] There are two primary strands within state action doctrine jurisprudence. The first strand involves circumstances in which a private actor performs a traditional, exclusive public function. The Court holds that it is not enough that the federal, state, or local government exercised the function in the past, still does, or that the function serves the public good the public interest in some way.[144] Rather, the government must have traditionally *and* exclusively performed the particular function.[145] Additionally, to qualify as a state action, private conduct must not only comprise something that the government traditionally does, but something that *only* the government traditionally does.[146] The Court has stressed that 'very few' functions fall into the first category, including running elections and operating a company town.[147]

---

[139] *Denver Area Ed Telecomm Consortium v FCC*, 518 US 727, 737 (1996).
[140] *Edmonson v Leesville Concrete Co*, 500 US 614, 619 (1991).
[141] *Manhattan* (n 137) 1928.
[142] *New York Times Co v Sullivan*, 376 US 254, 265 (1964).
[143] See Erwin Chemerinsky, 'Rethinking State Action' (1985) 80 Northwest U L Rev 503.
[144] See *Rendell-Baker v Kohn*, 457 US 830, 849 (1982).
[145] ibid.
[146] *Edmonson* (n 140) 640.
[147] See *Flagg Bros, Inc v Brooks*, 436 US 149, 158 (1978). Examples of functions that *do not* fall within this category include running sports associations and leagues, administering insurance payments, operating nursing

The second strand involves circumstances in which the government becomes entangled with private parties including, inter alia, 'mak[ing] extensive use of state procedures with the overt, significant assistance of state officials', when the injury caused to a private actor is 'aggravated in a unique way by the incidents of governmental authority',[148] when the government compels the private entity to take a particular action, or when the government acts jointly with the private entity.[149] The USSC has applied this strand when race or gender is used in the exercise of peremptory challenges in private civil litigation,[150] when a creditor obtains a writ of prejudgment attachment from a court,[151] and when the state uses its power to award damages for libel in actions brought by public officials against critics of their official conduct.[152]

The USSC's articulation and application of the state action doctrine has faced sustained and pointed criticism from constitutional scholars for both its lack of clarity and inconsistent application. Such criticism extends as far back as the 1950s, when perhaps no other subject attracted more attention in American law review articles.[153] In 1967, Yale Law Professor Charles Black described the doctrine as a 'conceptual disaster area' and a 'torchless search for a way out of a damp echoing cave'.[154] According to many contemporary constitutional scholars the last fifty years of USSC jurisprudence has not improved matters, and the doctrine

homes, providing special education, representing indigent criminal defendants, resolving private disputes, and supplying electricity. See *Manhattan* (n 137) 1928. Relatedly, the USSC has recognised that a private entity may, under certain circumstances, be deemed a state actor when the government has outsourced one of its constitutional obligations to that entity. For example, the Court had held that when a State contracts with private actors to provide medical care to prisoners, those private actors become state actors. Accordingly, any harm suffered by a prisoner from medical care by the State's decision to incarcerate the prisoner and put his medical care in the hands of private doctors. The fact that a doctor is a private contractor makes no difference because it is the actor's function that determines whether the conduct at issue could fairly be attributed to the State. The Court emphasised that if the rule were any different, a State would effectively be free to contract out all services that it is constitutionally obligated to provide, thereby leaving its citizens with no means for vindication of their fundamental rights. See *West v Atkins*, 487 US 42 (1988).

[148] See *Edmonson* (n 140) 615.

[149] ibid. For an in-depth discussion of the USSC's entanglement jurisprudence, see Chemerinsky (n 143).

[150] *Edmonson* (n 140) 631.

[151] *Lugar v Edmonson Oil Co*, 457 US 922 (1982).

[152] *Sullivan* (n 142) 283.

[153] Chemerinsky (n 143) 503.

[154] Charles L Black, Jr., 'Forward: "State Action," Equal Protection, and California's Proposition 14' (1967) 81 Harv L Rev 69, 95.

remains 'notoriously confusing, if not incoherent'.[155] Even the Court has acknowledged that its cases applying the state action doctrine 'have not been a model of consistency'.[156]

Another principle that merits attention in discussions regarding the application of fundamental rights that is often ignored in discussions of the state action doctrine is the principle of horizontality or horizontal effect. Unfamiliar to most Americans and relevant to a much greater degree in the European constitutional framework, as discussed below, the principle of horizontality denotes the application of certain fundamental rights to disputes by and between individuals, rather than to disputes between individuals and the state.[157] Assessing the operation of horizontality within constitutional law is an important question of constitutional construction, that is, how fundamental rights enter private relations within a particular legal framework.[158] This, in turn, may impact the extent to which these rights interact with other norms within a particular legal system.[159] Horizontality is particularly relevant to the discussions concerning intermediary liability in Chapters 4 and 6, which examine the implications and challenges of placing compulsory obligations on digital intermediaries to regulate the speech of users.

A useful definition of horizontal effect is provided by Eleni Frantziou in *The Horizontal Effect of Fundamental Rights in the European Union: A Constitutional Analysis*, in which she identifies three main types of horizontality: (1) the imposition of direct obligations on private parties (direct effect); (2) the indirect application of fundamental rights by the courts in disputes between private parties (indirect effect); and (3) the alteration of private relations by a right imposed on the state (state-mediated effect).[160] Frantziou identifies two subcategories of state-mediated effect: (1) where the state, through the public institution of the court, is conceptualised as taking part in all private proceedings and, as a result, is prevented from allowing private claims that breach fundamental rights; and (2) extensive positive obligations

[155] See Martha Minow, 'Alternatives to the State Action Doctrine in the Era of Privatisation, Mandatory Arbitration, and the Internet: Directing Law to Serve Human Needs,' (2017) 52 Harv Civil Rights-Civil Liberties L Rev 145. See also Mark Tushnet, 'The Issue of State Action/Horizontal Effect in Comparative Constitutional Law' (2003) 1 ICON 79, 80 (referencing Black's 1967 article and stating '[t]he situation has not improved since Black wrote'.).
[156] *Lebron v Nat'l RR Passenger Corp*, 513 US 374, 378 (1995).
[157] See Frantziou (n 134).
[158] ibid 1.
[159] ibid.
[160] ibid 38 - 39.

for the state, which entail its stepping in to actively protect fundamental rights against abuse by private actors.[161]

At first glance, the state action doctrine and horizontal effect appear to be contradictory principles, the former premised on the idea that fundamental rights may only be enforced against the acts of the government while the latter provides for the protection of fundamental rights in private relationships. However, many constitutional scholars argue that horizontality is present in the American constitutional framework, though, there is significant disagreement regarding to what extent and on what basis.[162] These arguments primarily focus on cases involving the 'entanglement' strand of the state action doctrine. The most oft cited case in support of the argument that horizontality is part of the American constitutional framework is *Shelley v. Kramer,* in which the USSC considered two cases involving the question of whether the Equal Protection Clause of the Fourteenth Amendment prohibited enforcement by state courts of restrictive covenants that excluded non-Caucasians from the ownership or occupancy of designated real property. [163] In both cases, the purchase of designated property by African-Americans spurred a request from other owners for judicial enforcement of the restrictions, which were contained in agreements between private actors. The actions of the state consisted only in the enforcement of these agreements. Thus, the crucial question was whether this distinction meant that the challenged actions constituted private rather than state conduct. If the answer to this question was yes, then the Fourteenth Amendment did not apply because the state action doctrine 'erects no shield against merely private conduct, however discriminatory or wrongful'.[164]

---

[161] ibid 39.

[162] See, e.g., Tushnet 'The Issue of State Action' (n 155) 79 - 80. Tushnet uses the terms 'state action' and 'horizontal effect' interchangeably, describing horizontal effect as the '"state action" problem' and frames the relevant question as 'the *way* in which the government is implicated in decisions by private employers and the like sufficient to place some duties on either the government or the private actors. This is the issue of state action in United States constitutional law, more commonly known elsewhere as the issue of horizontal effect of constitutional provisions'. Stephen Gardbaum argues that while private actors are not bound by constitutional rights in the United States they are indirectly subject to (and may be adversely affected by) them because such rights govern the laws that private actors invoke and rely on against one another. As a result, constitutional rights may either prevent such laws from protecting certain interests, choices, and actions of one private actor against another altogether, or place significant limits on their ability to do so. He argues that the extent of the reach of individual rights into the private sphere defies the standard understanding of the United States as creating a rigid public-private distinction in constitutional law, thereby epitomizing the vertical approach to this issue. Stephen Gardbaum, 'The "Horizontal Effect" of Constitutional Rights' (2003) 102 Mich L Rev 387.

[163] 334 US 1 (1948). Section 1 of the Fourteenth Amendment states, in relevant part, '[n]o state shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any state deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws'. US Const. amend. XIV.

[164] ibid 13.

The USSC concluded that the restrictive agreements 'standing alone' could not be regarded as a constitutional violation because so long as the enforcement was effectuated by voluntary adherence to their terms, there was no action by the state.[165] However, the Court determined that on the facts before it, 'there was more', i.e., that the discriminatory objectives of the agreements at issue were secured *only* by judicial enforcement by state courts of the restrictive terms.[166] For this reason, the Court concluded that there was 'no doubt that there has been state action in these cases in the full and complete sense of the phrase' and but for the active intervention of the state courts, the African-American property owners would have been free to occupy the properties in question without restraint.[167]

The USSC's holding in *Shelley* has proved to be extremely controversial, generated significant criticism for its lack of clarity, and reflects the broader inconsistencies in the Court's state action jurisprudence.[168] It suggests that, in certain and limited circumstances, the USSC may utilise the state action doctrine to render results that appear similar to the first form of state-mediated horizontal effect identified by Frantziou. In other words, there are a few limited circumstances in which the USSC will find state action and, thus, apply constitutional protections in cases between private actors. It is important to emphasise, however, that nowhere in its entanglement decisions does the USSC suggest that individuals bear the responsibility of respecting constitutional rights or that the state action doctrine fits within the broader international legal discourse concerning horizontality. Indeed, such terminology is entirely absent from the American constitutional lexicon and the contemporary USSC holds steadfast to the principle that the Constitution's protections do not extend to private conduct. Thus, the fundamental question in American constitutional law is whether the conduct at issue in a particular case, even if ostensibly undertaken by a private actor, may ultimately be attributable to the government. The landscape is different in Europe where, as discussed infra in Section B, horizontality takes centre stage in the discourse concerning the scope and application of fundamental rights.

[165] ibid.
[166] ibid 13-14.
[167] ibid 19.
[168] See, e.g., Johan van der Walt, *The Horizontal Effect Revolution and the Question of Sovereignty* (De Gruyter 2014).

2. The Public Forum Doctrine

In the United States, whether and to what extent free speech protections are triggered also depends on where speech occurs. For example, the USSC has long recognised that members of the public retain strong free speech rights on public streets and in parks, 'which have immemorially been held in trust for the use of the public and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions'.[169] With the concept of the 'traditional public forum' as a starting point, the USSC has recognised that members of the public have free speech rights on other types of government property and in certain other government programs that share essential attributes of a traditional public forum.[170] This 'forum based' approach is known as the public forum doctrine. To determine whether a public forum has been created for speech, courts look to the policy and practice of the government as well as the nature of the property and its compatibility with expressive activity to discern the government's intent.[171] In applying this doctrine, the Court divides public spaces into three categories: traditional public forms, designated public forums, and non-public forums.

Traditional public forums include those places 'which by long tradition or by government fiat have been devoted to assembly and debate,' such as parks, streets, and sidewalks.[172] Government proscriptions on speech in traditional public forums are subject to strict scrutiny. In such places, expressive activity will rarely be incompatible with the intended use of the property, as is evident from the facts, and that they are 'natural and proper places for dissemination of information and opinion'.[173] The second category of public property is the designated public forum, which is a non-public forum that the government has opened for all types of expressive activity.[174] Designated public forums are subject to the same limitations as those governing traditional public forums.[175] A subset of the designated public forum, the 'limited' public forum, exists 'where the government opens a non-public forum but limits the expressive activity to certain kinds of speakers or to the discussion of certain subjects'.[176] A

[169] *Perry Educ Ass'n v Perry Local Educators' Ass'n*, 460 US 37, 45 (1983) (internal quotations and citations omitted).
[170] ibid.
[171] See *Cornelius v NAACP*, 273 US 788, 802 (1985)
[172] *Perry* (n 169) 45.
[173] *Schneider v New Jersey*, 308 US 147, 151 (1939).
[174] See *Cornelius* (n 171) 802.
[175] ibid 830.
[176] See *NY Magazine v Metro Transp Auth*, 136 F3d 123, 128 (2nd Cir 1998).

limited public forum is created only where the government 'makes its property generally available to a certain class of speakers,' as opposed to reserving eligibility to select individuals who must first obtain permission to gain access.[177] In a limited public forum, it is not history or tradition, but the government's own acquiescence in the use of the property as a forum for expressive activity that makes it compatible with the uses to which the place is normally put.[178]

When the state establishes a limited public forum, it is not required to, and does not, allow persons to engage in every type of speech, and may be justified in reserving its forum for certain groups or for the discussion of certain topics. Specifically, 'the necessities of confining a forum to the limited and legitimate purposes for which it was created may justify the state in reserving it for certain groups or for the discussion of certain topics'.[179] However, once it has opened a designated forum, a state must respect lawful boundaries it has itself set and may not exclude speech where its distinction is not reasonable in light of the purpose served by the forum, nor may it discriminate against speech on the basis of viewpoint.[180] Common examples of limited public forums include public university meeting facilities opened for student groups, open school board meetings, city-leased theatres, and subway platforms opened to charitable solicitations.[181]

Finally, non-public forums are government property that is not open for public communication by tradition or designation. Access to a non-public forum can be restricted if the restrictions are reasonable and are not intended to suppress expression merely because public officials oppose the speaker's views.[182] Government regulation of expression in non-public forums is subjected to a more lenient standard of scrutiny and need only be reasonable and viewpoint neutral. Examples of non-public forums include airport terminals, highway overpass fences, and interstate rest stop areas.[183]

B. Obligations to Respect Free Rights in Europe: The Emergence of Horizontal Effect

[177] *Arkansas Educ Television Comm'n v Forbes*, 523 US 666, 679 (1998).
[178] ibid.
[179] *Rosenberger v Rector and Visitors of University of Virginia*, 515 US 819, 829 (1995).
[180] ibid.
[181] *Hotel Employees & Restaurant Employees Union v New York Dep't of Parks & Recreation*, 311 F3d 534, 545 (2nd Cir 2002).
[182] *Cornelius* (n 171) 800.
[183] See *Center for Bio-Ethical Reform, Inc v Honolulu*, 455 F3d 910, 919 (9th Cir 2006).

Unlike in the United States, the principle of horizontal effect plays a significant role in the European constitutional framework and is a popular topic in contemporary European constitutional discourse. However, both the meaning and proper scope of horizontality remain contested within the multi-layered framework of human rights protection on the continent.[184] The complexities and nuances of horizontal effect, and the debates relating thereto, fall outside of the scope of this thesis. Nor is my objective here normative, that is, to champion one conception or application of horizontal effect over another. Rather, my purpose is to explicate the general principles underlying horizontal effect in order broadly sketch the current landscape of horizontality in Europe.

This Section considers the approaches of the EU and the ECtHR, which collectively represent the scope and interpretation of horizontality on the continent. In EU law, horizontality arises only in cases where state protection of fundamental rights in private relations through the legislative function has failed or is non-existent.[185] In other words, horizontality in the EU is predicated upon a failure of an EU Member State to ensure the effective enforcement of EU fundamental rights.[186] While the EU Charter envisages that fundamental rights are capable, at least in principle, of creating obligations between private parties, questions regarding the extent to which this is true, and in what areas of EU law horizontal effect operates, remain open questions.[187] In recent years, a series of cases concerning different fundamental rights, including religious freedom, data protection, and age discrimination have rekindled the debate over the proper scope of, and basis for, horizontality in the EU.[188] The legal systems of EU Member States appear overwhelmingly inclined to choose indirect effect.[189]

With respect to the application of the ECHR, it is generally uncontested that ECHR rights generate some level of horizontal effect at the national level and that private parties may

[184] Frantziou (n 134) 37.
[185] ibid.
[186] ibid. As discussed in the Introduction, the EU Charter instructs that the fundamental rights recognised therein are to be given the same meaning and scope as ECHR rights.
[187] ibid.
[188] ibid. For additional scholarly commentary regarding horizontality in the contemporary EU, see Dorota Leczykiewicz, 'The Judgment in Bauer and the Effect of the EU Charter of Fundamental Rights in Horizontal Situations' (2020) 16 ERCL 323; Aurelia Colombi Ciacchi, '*Egenberger* and Comparative Law: A Victory of the Direct Horizontal Effect of Fundamental Rights' (2018) 5 EJCL 207; Cian C Murphy, 'Using the EU Charter of Fundamental Rights Against Private Parties after *Association De Médiation Sociale*' (2014) 2 EHRLR 170.
[189] ibid (Frantziou) 38.

adversely affect an individual's enjoyment of at least certain rights.[190] The establishment and development of horizontal effect by the ECtHR is grounded in a theory of positive obligations on the state to protect the enjoyment of fundamental rights in the context of relations between individuals or in conflicts between private parties and competing fundamental rights.[191] While not generally translated into distinct legal rights, positive obligations tend to be used as important arguments in Article 10 ECHR case law.[192]

In the context of Article 10, the ECtHR has made clear that freedom of expression is to be respected by governments, parliaments, and the judicial authorities of Member States.[193] In several cases, the ECtHR has opined on how Article 10 ECHR should be applied in private legal relationships and has adjudicated interferences by private persons in the light of Article 10(2). While the ECHR envisages horizontal effect through the imposition of protective duties, beyond that, the Council of Europe depends on Member States to determine whether and to what extent they may allow further avenues for horizontality.[194] In some countries the ECHR is given precedence over national law and its provisions have direct effect;[195] in other countries the ECHR has been 'indirectly' incorporated into domestic law.[196] Rather than adopting a general theory of horizontality, the ECtHR employs a case-by-case assessment of the potential for positive obligations with respect to particular rights.[197] The recognition by the ECtHR of a horizontal effect of Article 10 and of the positive obligations for Member States to protect the right to freedom of expression has further extended the scope of the right to freedom of expression in Europe, regardless of how precisely the ECHR is internally applied or guaranteed in Member States.[198]

---

[190] Gavin Phillipson and Alexander Williams, 'Horizontal effect and the constitutional constraint' (2011) 74(6) MLR 878.
[191] Jean-François Akandji-Kombe, 'Positive Obligations under the European Convention on Human Rights: A Guide to the implementation of the European Convention on Human Rights (Human Rights Handbooks No 7)' (*Council of Europe*, 2007).
[192] Brittan Heller and others, 'Freedom of Expression: A Comparative Summary of United States and European Law' (*Transatlantic Working Group*, 3 May 2019) 10 - 11.
[193] Dirk Voorhoof, 'The European Convention on Human Rights: The Right to Freedom of Expression and Information Restricted by Duties and Responsibilities in a Democratic Society' (2015) 7 JOUR 1.
[194] ibid 3. See, e.g., *Frasila and Ciocirlan* v *Romania*, App no 25329/03 (ECtHR, 10 May 2012).
[195] For example, Article 91 of Poland's Constitution states that 'a ratified international agreement shall constitute part of the domestic legal order and shall be applied directly, unless its application depends on the enactment of a statute. See 'Report of the Implementation of International Human Rights Treaties in Domestic Law and the Role of the Courts' (*Venice Commissio*n, 8 December 2014) CDL-AD(2014)036.
[196] For example, in the UK the ECHR is incorporated into domestic law by way of indirect horizontal effect. See generally Edward P Bates and others, *Law of the European Convention on Human Rights* (OUP 2009).
[197] Phillipson and Williams (n 180) 883.
[198] See Voorhoof (n 183) 37.

Advances in technology and the proliferation of speech on social media platforms raise significant questions regarding the scope and interpretation of the state action and public forum doctrines in the United States as well as the scope and impact of horizontality more generally in the digital age. These questions are examined in more detail in Part II, particularly in Chapters 5 and 6.

Conclusion

This chapter examines the doctrinal underpinnings of freedom of expression in the United States and Europe with respect to the threshold questions of what constitutes speech within constitutional frameworks, what classes of speech fall outside of such protections, and who is obligated to respect free speech rights. In addition to providing a doctrinal foundation for the debates over extreme speech explored in Part II, this examination leads to a richer and more contextualised understanding of each approach. Thus, what broader observations may be gleaned from this examination?

First, while an inquiry into the protections afforded to expressive conduct reveals a consensus in the USSC and the ECtHR that the right to freedom of expression extends beyond traditional modes of expression, the different tests applied in expressive conduct cases reflect significant divergences in doctrinal approaches. For example, while the USSC applies distinct tests in cases involving conduct and traditional modes of expression, the ECtHR applies a balancing test to all cases it determines fall within the ambit of Article 10, regardless of the type of expression at issue. Additionally, the inherent scepticism on content and viewpoint-based restrictions on expression in the United States is largely absent from the European conception of free speech, which encourages and, in certain circumstances, requires such restrictions.[199] An appreciation of these differences is imperative for a full understanding of the arguments made, recommendations offered, and conclusions reached in Part II.

Second, the doctrinal exceptions to free speech protections highlight the considerable challenges placed upon courts that adjudicate disputes over free speech proscriptions within broader constitutional frameworks that provide expansive protections to, and emphasise the importance of, freedom of expression. These challenges are apparent in the jurisprudence

[199] This is not to suggest that the European approach does not place a high value on speech in public discourse. See, e.g., *Dichand* (n 29).

relating to exceptions to free speech protections. The USSC jurisprudence in this area reveals a degree of judicial discord regarding the normative basis for excluding certain categories of speech from the protection of the First Amendment as well as inconsistencies within majority opinions regarding why and to what extent such speech warrants constitutional scrutiny. The ECtHR faces similar challenges in delineating exceptions to the broad protections of speech articulated in Article 10. Its application of the margin of appreciation and Article 17 in free speech cases has resulted in methodological inconsistencies, and its jurisprudence lacks a coherent test for deciding what classes of extreme speech fall outside of Article 10. These issues are particularly relevant to the discussions of 'hate speech' in Chapters 3 and 6.

Third, doctrinal principles concerning what actors must respect free speech rights reflect broader differences regarding how fundamental rights are conceptualised within legal frameworks. For example, the state action and public forum doctrines reflect, for the most part, an enduring tradition of negative freedom in the United States and an emphasis on preserving a sphere of 'liberty' for each individual to express his or her views without government constraints. The way in which horizontal effect is understood in Europe, in contrast, grounds the application of fundamental rights within a broader conception of rights holders, that is, that individuals hold rights not just in relation to the government but, in certain circumstances, in relation to one another. However, cases like *Shelley* suggest that horizontality is present in the American approach, though it is not often recognised and the legal discourse regarding the state action doctrine largely ignores the broader horizontality debates that are occurring in Europe. An understanding of the contemporary European discourse concerning horizontality and recent debates among American judges and scholars regarding the proper scope of the state action and public forum doctrines in the digital age reveal the degree to which different jurisdictions are grappling with similar questions resulting from the ubiquity of expression on online platforms owned by private actors. These issues inform Part II and are explored in detail in Chapter 6 with respect to increasingly aggressive efforts at the nation and supranational levels to influence in the ways in which digital intermediaries moderate the speech of users.

Finally, this examination touches on the interconnectedness between legal doctrine and legal philosophy. A firm understanding of both doctrine and philosophy are required for a CCL inquiry of this nature. Accordingly, Chapter 2 provides the philosophical grounding for this thesis by examining the most relevant philosophical divergences in the European and American

approaches to the regulation of extreme speech that inform the examinations undertaken in Part II.

Chapter 2

Why Do We Protect Speech? How Legal Philosophy Aids in Understanding the Key Differences in the Protections Afforded to Extreme Speech in Europe and the United States

---

'It is impossible to draw a sharp line between legal and philosophical principles arguments in this context. The literal approach to textual interpretation, of some serviceable use in the construction of detailed statutes, is of little assistance in elucidating the meaning of freedom of speech (or freedom of expression) provisions which are invariably framed in broad, general terms.'[1]

Introduction

In order to develop a full and precise understanding of the differences in the European and American approaches to regulating extreme speech and how these differences inform the discussions in Part II, we must look beyond the four corners of the relevant human rights instruments examined in Chapter 1 to the philosophical principles that govern freedom of expression in European and American constitutional frameworks, in particular, those that illuminate the marked differences in the protection afforded to extreme speech. For any principled and persuasive argument concerning the proper limits on the right to freedom of expression must contend with disagreements about why this right is worthy of protection in the first place.

Constitutional theories of free speech are underpinned by two features - theories of constitutional interpretation and an interpretative methodology that delineates how to treat different types of speech and on what basis.[2] This thesis is concerned with the second feature.

[1] Eric Barendt, *Freedom of Speech* (OUP 2005) 2.
[2] See C Edwin Baker, 'Autonomy and Hate Speech' in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy* (OUP 2010) 259.

As discussed in Chapter 1, the text of human rights instruments provides part of the overall picture of what free speech means in a particular place. The interpretation of those texts by courts is also important, not only because it tells us what 'speech' is, but because it enriches our understanding of the connection between how the right to freedom of expression is conceptualised in a given framework and the philosophical foundations of the right itself.[3]

In normative terms, legal philosophy tells us what the law should be, for example, what types of speech warrant protection from government inference and why.[4] Thus, philosophical arguments are relevant to constitutional interpretation because they assist courts in contextualising the right to freedom of expression within a particular legal system.[5] Legal philosophy also helps to explain why free speech is regarded differently in different systems; how we approach debates over free speech is largely influenced by the ways in which this right is conceptualised in a particular place. Therefore, it is important to consider the extent to which broader human rights principles underly specific rights within a particular system, and the ways in which courts are influenced by the broader philosophical principles that underpin the systems in which they are situated.[6]

It is important at this juncture to reiterate that this thesis does not endeavour to make a contribution to the voluminous literature concerning the philosophical underpinnings of freedom of speech in Europe and the United States. Rather, the objective here is to identify and examine the most relevant philosophical divergences in the European and American approaches to the regulation of extreme speech that inform the examinations of extreme speech undertaken in Part II. The principles discussed in this chapter provide the philosophical grounding for Chapters 3 through 6 by exploring the ways in which the work of legal philosophers enriches our understanding of the European and American approaches to 'hate speech', terrorist-related expression and disinformation, as well as why relevant divergences exist. They also aid in evaluating the strengths and weaknesses of each approach.

With the foregoing considerations in mind, this chapter proceeds as follows. Part I examines theories of democratic legitimacy and how these aid in appreciating the marked differences in

[3] See Robert Post, 'Participatory Democracy and Free Speech', (2011) 97(3) Va L Rev 477, 487.
[4] See Dieter Grimm, 'Freedom of Speech in a Globalized World' in Hare and Weinstein (n 2) 12 - 13.
[5] James Weinstein, *Hate Speech, Pornography, and Radical Attacks on Free Speech Doctrine* (Routledge 1999).
[6] Barendt (n 1) 5 - 6.

the value afforded to public discourse in the European and American approaches to speech regulation as reflected in extreme speech jurisprudence. This analysis highlights the extent to which the American emphasis on the importance of debate on matters of public concern is linked to a particular conception of democracy that is predicated on the principle that citizen participation in public discourse must be free and unfettered. This analysis is particularly illuminating with regard to the differences in the American and European approaches to the regulation of 'terrorist-related expression examined in Chapter 4, including the European tolerance of glorification related offences. Part II examines how differing conceptions of the role and relevance of dignity in European and American free speech frameworks reflect distinct philosophical conceptions of autonomy and analyses how these differences manifest in relevant case law. This discussion aids in understanding the marked differences in the European and American approaches to the regulation of 'hate speech' discussed in Chapter 3. Part III examines how theories of liberty and equality aid in understanding why the practice of attaching duties and responsibilities to the exercise of rights is common throughout Europe and expressly rejected in the United States. This examination grounds the discussion of the scope of immunities granted to state actors in Europe and the United States in Chapter 5, as well as the criticism of the American reliance on theories of democratic legitimacy proffered in Chapter 6.

I. Theories of Democratic Legitimacy and the Value Afforded to Public Discourse in European and American Free Speech Frameworks

As discussed in Chapter 1, while expression that is characterised as contributing to public discourse enjoys a privileged status in both European and American constitutional frameworks, such expression receives far greater protection in the United States. This reflects fundamental differences regarding the value afforded to public discourse and whether it is the province of the government to determine whether certain types of speech warrant inclusion in public debate. This section examines the ways in which theories of democratic legitimacy assist in accounting for the significant doctrinal divergences between Europe and the United States in this area.

A. Free Speech Theories Rooted in Democratic Legitimacy

Whether deontological, descriptive, or normative, theories of democratic legitimacy are grounded in notions regarding the appropriate relationship between the individual and the state and are premised on the principle that allowing citizens the opportunity to meaningfully participate in public discourse is a necessary precondition for a legitimate democracy.[7] These are distinct from consequentialist arguments such as those advanced by critical race theorists, which are concerned with the harm to marginalised and vulnerable groups that may result from extreme speech.[8] This is not to suggest that those advancing theories of democratic legitimacy are unconcerned with social injustice or systemic inequality but, rather, they view such concerns as illegitimate grounds for governmental restrictions on extreme speech due to the deleterious effect of such restrictions on democratic legitimacy.

Much of the contemporary scholarship concerning the relationship between extreme speech and democratic legitimacy in the United States and Western Europe builds on the seminal work of the late Ronald Dworkin, for whom freedom of expression was a 'basic right' for reasons of 'basic principle'.[9] The crux of Dworkin's theory is that allowing all citizens the opportunity to meaningfully participate in public discourse is a necessary precondition for democratic legitimacy.[10] Accordingly, it is illegitimate for the government to impose a collective or official decision on dissenting individuals using the coercive powers of the state unless that decision respects each individual's status as a free and equal member of the community.[11] Relatedly, 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

[7] Amanda R Greene and Robert M Simpson (2017) 'Tolerating Hate in the Name of Democracy' 80(4) MLR 746.
[8] See, e.g., Mari Matsuda, *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Westview Press 1993).
[9] Ronald Dworkin, 'A New Map of Censorship' (*Index on Censorship*, May/June 1994). Of course, Dworkin was not the first philosopher to link meaningful participation in the political process with democratic legitimacy. For example, in the mid-twentieth century, Alexander Meiklejohn posited that the right to express disagreements with laws was a morally and politically necessary condition of compelling people to obey laws with which they disagreed and opined that people would be more inclined to obey those laws when they were given the opportunity to object than would be the case were their dissenting voices to be stifled by state action. See Alexander Meiklejohn, *Free Speech and its Relation to Self-Government* (Harper & Brothers Publishers 1948) 10 - 11, and Alexander Meiklejohn, *Political Freedom: The Constitutional Powers of the People* (Harper & Brothers Publishers 1960) 100. See also Immanuel Kant, *Groundwork of the Metaphysics of Morals* (1979) (Mary Gregor tr, 1st edn, 1998) and John Locke, *Second Treatise of Government* (1689) (CB Macpherson ed, 1st edn, 1980), both of which are discussed by James Weinstein in 'Hate Speech Bans, Democracy, and Political Legitimacy' (2017) 32 Const Comment 527.
[10] Ronald Dworkin, 'Even Bigots and Holocaust Deniers Must Have Their Say' *The Guardian* (14 February 2006) <http://www.theguardian.com/world/2006/feb/14/muhammadcartoons.comment> accessed 10 March 2019.
[11] ibid.

just in the hope of influencing others...but also just to confirm his or her standing as a responsible agent in, rather than a passive victim of, collective action'.[12]

In Dworkin's view, we must be willing to tolerate the expression, however insulting or offensive, of those who oppose laws because only a community that allows such expression may legitimately adopt such laws.[13] Put another way, Dworkin reasons that we can only expect bigots to accept the verdict of the majority once the majority has spoken, so we must permit them to express their bigotry in the process whose verdict we expect them to respect.[14] While we may disagree with their opinions and believe them to be false, a government that claims the power to identify and impose truth is not fully legitimate.[15] While Dworkin emphasises that we must protect minority groups from the damaging consequences of bigotry, he argues that we must do so by passing laws that target discrimination and other ills, not by prohibiting expression of any of the attitudes that we believe contribute to inequality.[16] For if we intervene at the point at which opinions are formed, 'we spoil the only democratic justification we have for insisting that everyone obey these laws, even those who hate and resent them'.[17] Because Dworkin's theory is grounded in the concept of negative liberty, his conception of a 'fair opportunity' does not include any obligation of the part of the government to take affirmative steps to create optimal conditions for citizen participation in public discourse.

Other prominent theories concerning the relationship between free speech and democratic legitimacy include those of American free speech scholars Robert Post and James Weinstein. Post is primarily concerned with descriptive legitimacy and the ways in which freedom of expression underwrites democratic legitimacy by allowing persons to participate in public discourse when they believe that the state will generally be responsive to public opinion.[18] Persons prevented from expressing opposition to government acts are likely to regard such acts as unfair. For Post, the purpose of free speech is political rather than epistemological because

---

[12] ibid.
[13] ibid.
[14] ibid. For a related argument concerning the relationship between free speech and government legitimacy, see Kenan Malik, 'Interview with Kenan Malik' in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulations and Responses* (CUP 2012) 81-91.
[15] Ronald Dworkin, 'Reply to Jeremy Waldron' in ibid (Herz and Molnar) 343.
[16] ibid. See also Nadine Strossen, 'Interview with Nadine Strossen,' in Herz and Molnar (n 14) 381-382 (In lieu of restrictions on extreme speech, which she argues fail to address the root causes underlying this type of speech, Strossen advocates for addressing discrimination and violence against minorities in an effort to ensure that all individuals have security in terms of basic safety, economic needs, educational opportunity, and job opportunity.).
[17] ibid (Dworkin).
[18] Robert Post, 'Legitimacy and Hate Speech' (2017) 32 Const Comment 651, 656.

it represents a guarantee of political equality in the formation of public opinion.[19] Speech is worthy of protection because it reinforces systemic legitimacy by providing persons with widely divergent opinions the opportunity to experience as legitimate a government that acts in ways that are inconsistent with their own ideas.[20] Democratic legitimacy increases the likelihood of a viable political society in which persons with different views live peaceably under a single system of government. Because a viable political society is imperative to the maintenance of values, Post argues that descriptive legitimacy is among the most fundamental of democratic principles.[21]

While there is much overlap between the theories of Post and Weinstein, the latter is more concerned with normative rather than descriptive legitimacy. Weinstein's theory is grounded in notions of equality, the crux of which is that, generally speaking each individual in society is of equal moral worth and, as a result, is entitled to have his or her views treated with equal respect by the government.[22] In Weinstein's view, there comes a point at which a speech restriction, like selective disenfranchisement, can so profoundly disrespect both the interests and equal moral worth of some individuals that the restriction can have an effect not just on the legitimacy of the legal system, which is Post's primary concern, but also on particular laws enforced against those whose ability to oppose these laws was severely curtailed.[23] For Weinstein, if an individual is excluded from participating in public discourse because the government disagrees with his or her views, or because it finds the ideas expressed too disturbing or offensive, any decision taken as a result of that decision would, with respect to that citizen, lack legitimacy.[24]

A more recent contribution to the scholarship concerning free speech and democratic legitimacy comes from Eric Heinze.[25] Building on the work of Dworkin, Post, Weinstein and others, Heinze proposes a theory grounded in the notion that expression in public discourse is not just an important right within democracy, to be balanced and measured against other rights,

[19] ibid.
[20] ibid.
[21] ibid.
[22] James Weinstein, 'Viewpoint Discrimination, Hate Speech, and Political Legitimacy: A Reply' (2017) 32 Const Comment 715, 778 - 779.
[23] ibid 727.
[24] James Weinstein, 'Participatory Democracy as the Central Value of American Free Speech Doctrine' (2011) 97 Va L Rev 491, 498.
[25] Eric Heinze, *Hate Speech and Democratic Citizenship* (OUP 2016).

but is materially constitutive of democracy itself.[26] Specifically, Heinze posits that democracy's legitimating expressive conditions derive from the 'citizen's prerogative of non-viewpoint punitive expression within public discourse'.[27] It is the prerogative of non-viewpoint-punitive expression within public discourse, along with any necessary derivatives of that prerogative, which legitimates states as democracies.[28] Heinze considers this attribute of democracy as something citizens 'carry around with us always and everywhere within the borders of our democracy and which cannot be regulated for the sake of democracy' for the reason that is 'signally constitutes democracy' and is its 'defining element'.[29] Heinze distinguishes his theory from those of Dworkin and others who anchor the normative criteria of democratic legitimacy in individual rights rather than in democratic citizenship.[30]

B. The Relationship Between Theories of Democratic Legitimacy and the Value Afforded to Public Discourse

The extent to which the American approach to free speech is rooted in theories of democratic legitimacy is evident from the value afforded to public discourse by the USSC, which regards 'unencumbered public debate' as the 'essence of self-government'[31], and considers the right to receive information and ideas, regardless of their social worth, as fundamental to a free society.[32] In the United States, the freedom to speak one's mind is considered both an aspect of individual liberty, and thus a good unto itself, and as essential to the common vitality of society as a whole.[33] This reflects the profound American commitment that debate on public issues should be 'uninhibited, robust, and wide-open'[34] and to the principle that the government

[26] ibid 5. For a comprehensive critique of Heinze's theory, see Greene and Simpson (n 7).
[27] By 'legitimating expressive conditions, Heinze means those 'free speech conditions without which democratic processes cannot be said to be fully in place'. ibid 17.
[28] ibid 46.
[29] ibid 48 - 49.
[30] ibid 44. In arguing for viewpoint absolutism in the sphere of public discourse, Heinze limits his theory to longstanding, stable and prosperous democracies (LSPDs). While fully-fledged democracies may take a host of measures to eliminate discrimination, he argues that they cannot legitimately penalise citizens who enter the public sphere to oppose pluralist values, however provocatively they may do so. For Heinze, the citizen's prerogative of expression within public discourse remains necessary to democracy, not as a matter of sheer expedience on a consequentialist criterion, but as a matter of principle, i.e., 'on a deontological criterion'. ibid.
[31] *Garrison v Louisiana*, 379 US 64, 74-75 (1964).
[32] *Stanley v Georgia*, 394 US 557, 564 (1969). See also *Hustler Magazine, Inc v Falwell*, 485 US 46, 50 (1988) (opining that '[a]t the heart of the First Amendment is the recognition of the fundamental importance of the free flow of ideas and opinions on matters of public interest and concern'). James Weinstein astutely observes that, in descriptive terms, no other theory provides nearly as good an explanation of the USSC's free speech jurisprudence. See Weinstein 'Participatory Democracy' (n 24) 514.
[33] *Bose Corp v Consumers Union of United States,* 466 US 485, 503 - 504 (1949).
[34] ibid (quoting *New York Times Co v Sullivan,* 376 US 254, 270 (1964)).

has no authority to limit expression based on 'its message, its ideas, its subject matter, or its content'[35] According to the Court, these safeguards were created to 'assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people.'[36] Thus, freedom of expression protects the 'opportunity for free political discussion to the end that government may be responsive to the will of the people and that changes may be obtained by lawful means,' which is 'essential to the security of the Republic.'[37]

It is for these reasons that speech in public discourse occupies the highest rung on the hierarchy of First Amendment values and that content and viewpoint-based proscriptions on speech are presumptively unconstitutional.[38] The distinction between matters of private and public concern is therefore crucial, with the latter receiving more constitutional protection based on the underlying presumption that restricting speech on purely private matters does not implicate the same constitutional concerns.[39] While the USSC recognises that speech 'can stir people to action, move them to tears of both joy and sorrow, and…inflict great pain', it prohibits the government from reacting to that pain by punishing speakers because the United States has 'chosen a different course - to protect even hurtful speech on public issues to ensure that we do not stifle public debate'.[40] Moreover, while the exercise of the right to freedom of expression may manifest as 'verbal tumult, discord, and even offensive utterances,' the Court regards these as necessary side effects of the more general values of freedom that the process of open debate is meant to achieve.[41] Thus, even hurtful speech on public issues is protected to ensure that the government does not encumber public debate.[42]

While the USSC does not go so far as Heinze in conceptualising free speech as constitutive of

---

[35] See *Ashcroft v ACLU*, 535 US 564, 573 (2002).
[36] *Roth v United States*, 354 US 476, 484 (1957).
[37] *Stromberg v California*, 283 US 359, 369 (1931).
[38] See *Reed v Town of Gilbert*, 576 US 155, 162 (2015).
[39] *Snyder v Phelps*, 562 US 443, 452 (2011) (holding that restricting speech on purely private matters does not present a 'threat to the free and robust debate of public issues; there is no potential interference with a meaningful dialogue of ideas; and the threat of liability does not pose the risk of a reaction of self-censorship on matters of public import'.) (internal quotations and citations omitted). See also discussion of the meaning of 'public discourse' in the Introduction, p 4.
[40] ibid (*Snyder*).
[41] *Cohen v California* 403 US 15, 25 (in which the USSC opined that '[t]hat the air may at times seem filled with verbal cacophony is, in this sense not a sign of weakness but of strength. We cannot lose sight of the fact that, in what otherwise might seem a trifling and annoying instance of individual distasteful abuse of a privilege, these fundamental societal values are truly implicated. That is why wholly neutral futilities come under the protection of free speech as fully as do Keats' poems or Donne's sermons' and it is 'often true that one man's vulgarity is another man's lyric'.) (internal quotations omitted).
[42] See, e.g., *Snyder* (n 39) 460 - 461.

democracy, its jurisprudence reflects the general principles espoused in theories linking free speech with democratic legitimacy. Indeed, the Court's emphasis on unencumbered public discourse as a necessary precondition of ensuring that the government is responsive to the will of the people and that changes may be obtained through lawful means, demonstrate the extent to which democratic legitimacy functions as the primary philosophical tenet of the American free speech framework. This aids in contextualising relevant doctrine, including the emphasis on the value of public debate, the rule that viewpoint and content-based restrictions on expression are presumptively unconstitutional, and the Court's refusal to regulate extreme speech on the basis of the deleterious effects that may result, both in terms of the harm to the targets of such speech as well as to society more broadly.[43]

While the principle that speech in public discourse is worthy of protection is also present in the European free speech framework, it is one of several principles rather than the predominant one. For example, while the ECtHR opines that freedom of expression 'constitutes one of the essential foundations of a democratic society'[44] and that 'freedom of political debate is at the very core of the concept of democratic society',[45] the Council of Europe expressly eschews 'absolute liberalism' in favour of an approach predicated on the notion that not all ideas are deserving of circulation.[46] Moreover, notwithstanding the value it attributes political debate in democracies, the ECtHR opines that it is legitimate to expect from every member of a democratic society to avoid as far as possible expression that is gratuitously offensive to others.[47] This contrasts starkly with the USSC's decree that the government may not proscribe speech based on distaste for the way in which it is expressed.[48]

Additionally, as discussed in Chapter 1, the ECtHR balances the right to freedom of expression against other rights enumerated in the Convention, which are considered of equal value. For this reason, unlike the USSC, the ECtHR assesses the *value* of the speech based on its content

[43] The exception to this general rule is the *Brandenburg* test, which holds that the government may prohibit advocacy of the use of force or law violation only where 'such advocacy is directed to inciting or producing imminent lawless action and is likely to incite for produce such action'. *Brandenburg v Ohio*, 395 US 444, 447 (1969).
[44] *Maguire v United Kingdom,* App no 58060/13 (ECtHR, 3 March 2015) para 48.
[45] *Lingens v Austria,* App no 9815/82 (ECtHR, 8 July 1986) para 42.
[46] 'Report on the Relationship between Freedom of Expression and Freedom of Religion: the Issue of Regulation and Prosecution of Blasphemy, Religious Insult and Incitement to Religious Hatred (*Venice Commission*, October 2008).
[47] See, e.g., *Maguire* (n 44) para 49.
[48] See *Cohen* (n 41) 25.

to determine whether it adds any measurable value to public debate. In so doing, it permits Member States to serve as the arbiters of whether a certain type of speech makes a sufficient contribution to warrant inclusion in public debate. This aids in understanding why the types of content and viewpoint-based restrictions that are prohibited in the United States are commonplace throughout Europe. For example, as discussed in detail in Chapter 4, the ECtHR routinely upholds proscriptions on expression that target support for terrorism in the form of speech glorifying terrorist violence.[49]

Thus, while American and European free speech frameworks value the role of public debate in democratic societies, the American approach places much greater importance on this value over others due to its firm grounding in theories linking free speech to democratic legitimacy. However, recent developments in the United States, including the online disinformation campaign led by former President Trump and other prominent Republicans undermining the integrity of the 2020 election, suggest that the relationship between unfettered public debate and democratic legitimacy warrants re-examining in the digital age. Chapter 6 interrogates free speech theories grounded in democratic legitimacy following the insurrection at the United States Capitol following the 2020 election.

II. Theories of Autonomy and Conceptions of Dignity in European and American Free Speech Frameworks

Free speech debates surrounding the role of dignity in assessing the legitimacy of government proscriptions on extreme speech in the realm of public discourse are primarily grounded in theories of personal autonomy. Dignity, which often functions as a background principle that informs constitutional interpretation within a given human rights framework, may be best understood as a concept with a particular legal and political meaning that is dependent on the framework in which it is situated.[50] Generally speaking, as a legal concept, dignity aims to protect the innate supreme and inalienable value of human beings.[51] This value recognises human beings as ends-in-themselves. The normative result of this recognition is the protection of the subject status of human beings, i.e., their ability to become the authors of their lives and,

[49] See, e.g., *Gül and others v Turkey*, App no 4870/02 (ECtHR, 8 September 2010). See also Eliza Bechtold and Gavin Phillipson, 'Glorification of Censorship? Anti-Terror Law, Speech, and Online Regulation' in Adrienne Stone and Frederick Schauer (eds), *The Oxford Handbook on Freedom of Speech* (OUP 2020).
[50] Erin Daly, *Dignity Rights: Courts, Constitutions, and the Worth of the Human* Person (University of Pennsylvania Press 2013) 18 - 19.
[51] Matthias Mahlmann, 'Human Dignity and Autonomy in Modern Constitutional Orders' in Michel Rosenfeld (ed), *The Oxford Handbook of Comparative Constitutional Law* (OUP 2012) 389 - 390.

as a result, of their autonomy.[52] Put another way, the respect for dignity within human rights frameworks recognises a certain protection of the subject status of human beings as individuals, thereby implying the guarantee of their autonomy.[53]

While the notion of dignity plays a central role in both European and American free speech frameworks, it does so in conceptually distinct ways that influence the extent to which proscriptions on extreme speech are regarded as legitimate. The European conception of dignity is underpinned by a philosophical conception of autonomy that is rooted in the individual target of extreme speech. In contrast, the American conception of dignity is premised upon a philosophical conception of autonomy that is rooted in the individual speaker. These divergences, like those concerning the relationship between free speech and democratic legitimacy, reflect deeply entrenched philosophical disagreements concerning on what grounds governments may proscribe extreme speech. The following section provides an overview of three free speech theories rooted in autonomy that serve to elucidate the doctrinal differences discussed in Section 2. Differences in conceptions of autonomy and dignity in European and American free speech frameworks are particularly relevant to the debates over the regulation of 'hate speech' discussed in Chapter 3, including definitional challenges and questions regarding whether social harm is a legitimate basis upon which to restrict certain types of expression.

A. Free Speech Theories Rooted in Conceptions of Autonomy and Dignity

Jeremey Waldron and Steven Heyman advance free speech theories rooted in the autonomy and/or dignity of the targets of extreme expression.[54] Waldron is primarily concerned with the type of extreme speech commonly referred to as 'hate speech'. The crux of Waldron's theory can be distilled down to a few fundamental points. First, he believes that the value in restrictions on 'hate speech' derives from the protections they afford to minorities against the harms resulting from 'group libel'.[55] In Waldron's view, this term is preferable to 'hate speech' in that it recognises that speech attacking a group, such as 'Muslims Out' are reputational

[52] ibid 378
[53] ibid 379.
[54] Waldron is one of the preeminent proponents of regulating extreme speech and a staunch critic of the American approach.
[55] See, e.g., Jeremy Waldron, *The Harm in Hate Speech* (Harvard University Press 2012) 34 - 64.

attacks that represent assaults upon the dignity of the persons affected.[56]

Waldron uses dignity in the sense of these people's 'basic social standing, the basis of their recognition as social equals and as bearers of human rights and constitutional entitlements.'[57] He emphasises the importance of considering the way in which a person's status as a member of society in good standing is affirmed and sustained.[58] Waldron understands hate speech regulations as protecting a certain sort of important public good, 'a visible assurance offered by society to all of its members that they will not be subject to abuse, defamation, humiliation, discrimination, and violence on grounds of race, ethnicity, religion, gender, and in some cases sexual orientation'.[59] In his view, laws against group libel reflect 'a collective commitment on the part of society to uphold the fundamentals of people's reputations as members of society in good standing…vindicating the rudiments of their civic dignity as a necessary ingredient of public order.'[60]

Second, Waldron distinguishes between speech that is merely offensive and speech that amounts to group libel, that is, speech that represents an assault on people's dignity. Only the latter, in Waldron's view, merits regulation.[61] Waldron distinguishes between objective or social aspects of a person's standing in society, which are worthy of protection by the law, and subjective features, such as hurt feelings or anger, which are not.[62] This distinction is important to Waldron because it separates dignity from defamation and mere offense, even when that offense represents the core of what individuals believe is the identity of their group.[63] For Waldron, being disturbed by a shocking attack on one's views, even one's deepest and most cherished religious convictions, is not something people have a legitimate interest in being protected against.[64]

Like Waldron, Heyman's free speech theory is centred on the targets of extreme speech, though

[56] ibid 59.
[57] ibid.
[58] ibid 141-142.
[59] Jeremy Waldron, 'The Oliver Wendell Holmes Lectures, Dignity and Defamation: The Visibility of Hate' (2009) 123 Harv L Rev 1596, 1599.
[60] ibid 1600.
[61] Waldron '*Harm in Hate Speech'* (n 55) 122 - 123.
[62] Waldron 'Oliver Wendell Holmes Lectures' (n 59) 1613.
[63] ibid.
[64] ibid.

his theory also incorporates the rights of citizens more generally.[65] Drawing on the liberal natural rights tradition associated with John Locke and Immanuel Kant, Heyman advances a general theory of rights[66] founded on the duty of both states and individuals to respect the autonomy and dignity of human beings.[67] For Heyman, freedom of speech is understood within a broader framework of rights that must be exercised with due respect for the dignity and autonomy rights of others. 'Hate speech', which he defines as 'expression that abuses or degrades others on the basis of such characteristics as race, religion, and gender',[68] is not entitled to constitutional protection because it fails this test, thereby violating the principles that should govern democratic debate, including mutual respect among free and equal citizens.[69] Heyman disputes the USSC' characterisation of hate speech as lawful political speech, arguing that such expression 'transgresses the most basic ground rules for public discourse.'[70] Borrowing from Meiklejohn, Heyman argues that 'hate speech' is a 'form of abuse that violates the rules of order that make democratic deliberation possible'.[71]

Under Heyman's 'liberal humanist' approach, 'hate speech' infringes the right of individuals, chiefly, the 'the right to be recognised', as a human being, which he argues serves as the basis for all other rights.[72] Specifically, 'hate speech' violates the rights of its targets by, among other things, assaulting their dignity and denying their equal status as human beings and members of the community. As a result, 'hate speech' falls outside of a 'proper understanding of political debate'.[73] Rooted in respect for personhood, the right to recognition serves as 'the bond that constitutes the political community' and, as a result, individuals have a duty to recognise one another as human beings and citizens. 'Hate speech' warrants regulation because its aim is to 'dominate and subordinate others' and, in so doing, is inconsistent with those relations of mutual recognition in which each person can be expected to be respected by all as free and equal'.[74] Finally, like Waldron, Heyman emphasises the distinction between a

[65] Heyman offers what he characterises as an alternative theory of when speech should be protected under the First Amendment.
[66] Heyman also refers to his theory as 'liberal humanist'. See Steven Heyman, *Free Speech and Human Dignity* (Yale University Press 2008) 4.
[67] Steven Heyman, 'Hate Speech, Public Discourse, and the First Amendment' in Hare and Weinstein (n 2) 172.
[68] ibid 164.
[69] ibid 181. While Heyman is primarily concerned with the regulation of hate speech, his theory is equally applicable to the other forms of extreme speech discussed in this thesis.
[70] ibid 176
[71] ibid.
[72] ibid 170 - 171.
[73] ibid 179.
[74] ibid (internal quotations omitted).

violation of a person's rights and offensive expression, with only the former justifying legal regulation.[75]

In contrast to Waldron and Heyman, Edwin Baker proffers a theory of autonomy that is firmly rooted in the individual speaker.[76] A self-described 'advocate of almost absolute protection' of freedom of expression' and a proponent of America's strongly speech-protective approach, Baker's describes his theory as 'the best, though often unrecognized, explanation of existing American case law…'[77] Uninterested in instrumentalist justifications for free speech, such as the search for truth and substantive autonomy,[78] Baker's theory is grounded in a formal conception of speech autonomy, which respects a person's autonomy in her speech choices.[79] Baker's version of formal autonomy is premised on a person's authority or right to make decisions about herself.[80] His theory includes self-expressive rights such as 'a right to seek to persuade or unite or associate with others – or to offend, expose, condemn, or disassociate with them'.[81] In Baker's view, the state denies a person formal autonomy if the law denies her the right to 'use her own expression to embody her views'.[82] He argues that this formal conception of autonomy warrants 'virtually absolute protection from, and respect by, the state especially in relation to self-expressive or value-expressive behaviour'.[83] He starts from the premise that a legal order cannot reasonably claim legitimacy without, among other requirements, respecting people's autonomy.[84] Thus, the legitimacy of a legal order hinges, in part, on respecting the autonomy of the people obligated to obey its laws.[85] This respect is only achieved when people are permitted to express their own values, regardless of what those values are and irrespective of whether such expression causes harm to other people or makes government processes or the achievement of governmental aims more onerous.[86]

---

[75] ibid 181.
[76] Baker (n 2).
[77] ibid.
[78] Baker defines 'substantive autonomy' as involving 'a person's actual capacity and opportunities to lead the best, most meaningful, self-directed life possible'. ibid 143.
[79] C Edwin Baker, 'Autonomy and Free Speech' (2011) 27 Const Comment 251, 254.
[80] ibid 259
[81] Baker 'Autonomy and Hate Speech' (n 2) 142.
[82] ibid.
[83] Baker 'Autonomy and Free Speech' (n 79) 254.
[84] ibid 269.
[85] This is similar to Dworkin's theory, outlined above.
[86] Baker 'Autonomy and Hate Speech' (n 2) 142.

In the context of extreme speech, Baker conceives of relevant harm in a narrow sense, that is, harm to another's 'equal formal autonomy'.[87] Specifically, he posits that a speaker's formal autonomy may be limited only to the extent it harms another's equal formal autonomy.[88] He regards any restriction on 'hate speech' based on the injuries it may cause to another, including to their substantive autonomy, as generally impermissible. For this reason, he posits that prohibitions on 'hate speech' should generally be impermissible, even if arguably permissible in special, usually institutionally bound, contexts. These include where the speaker has no claimed right to act autonomously, such as in the employment context, where she has given up her autonomy in order to meet demands that are inconsistent with expressions of racism.[89] Baker rejects an emphasis on democratic foundations for free speech and adopts a basic premise of respect for individual's autonomy to which the law must conform. On this basis, the legal order must respect the autonomy even of the individual who would deny such respect to others in the community. In other words, 'the law must respect the freedom of the racist to express her views'.[90]

B. The Relationship Between Conceptions of Dignity and Theories of Autonomy in European and American Free Speech Frameworks

The theories of autonomy discussed above assist in elucidating the differing conceptions of dignity that underpin European and American free speech frameworks and how these differences manifest in doctrinal approaches to regulating extreme speech. While the term does not appear in the text of the United States Constitution, dignity plays a role in the USSC's First Amendment jurisprudence. The Court's application of dignity has faced criticism as fragmented and underdeveloped for failing to articulate exactly what the term denotes as well as its constitutive elements.[91] While neither as developed nor as extensive as that of the ECtHR discussed below, the treatment of dignity in the USSC's First Amendment jurisprudence is sufficiently developed to discern that its approach to extreme speech is informed by a theory of autonomy that, like Baker's, is firmly rooted in the individual speaker.

Thus, Baker's theory is particularly instructive here as it assists in understanding the USSC's

[87] Baker 'Autonomy and Free Speech' (n 79) 254.
[88] ibid.
[89] Baker 'Autonomy and Hate Speech' (n 2) 143.
[90] ibid.
[91] See, e.g., Aharon Barak, *Human Dignity: The Constitutional Value and the Constitutional Right* (CUP 2015) 206.

approach to the regulation of extreme speech. As discussed above, freedom of expression in the United States is designed and intended to remove governmental restraints from public discussion, leaving the decision as to what views are voiced largely in the hands of each citizen.[92] In the Court's view, 'no other approach would comport with the premise of *individual dignity* and choice upon which [the American] political system rests'.[93] The fierce protection of the right of self and value-expressive behaviour in the Court's free speech doctrine reflects a profound respect for the autonomy of all individuals, even those, as Baker notes, who would deny that same level of respect to others.

The Court regards injury to the dignity of the listener and/or targets of extreme speech as entirely irrelevant to whether such speech warrants First Amendment protection. Indeed, it expressly rejects a 'dignity' standard as applied to the targets of expression, reasoning that it is so inherently subjective that it would be inconsistent with the Court's 'longstanding refusal to [punish speech] because the speech in question may have an adverse emotional impact on the audience.'[94] While recognising that extreme speech may cause both social harm and harm to the targets, the American approach tolerates such harm because of the greater harm that may result from permitting the government to regulate speech based on content and viewpoint.[95] With respect to the effect that such speech may have on listeners, the Court consistently emphasises that citizens 'are often captives' outside the sanctuary of the home and subject to objectionable speech.'[96] As a result, the constitutionality of government efforts to proscribe speech solely to protect others from hearing it is dependent upon a showing that substantial *privacy* interests are being invaded in an essentially intolerable manner.[97] The USSC cautions that a broader view of government authority in this realm 'would effectively empower a majority to silence dissidents simply as a matter of personal predilections'.[98]

The Court's characterisation of freedom of expression as necessary for the development of individual dignity and its express disregard for the dignity of the targets of extreme speech reveal the extent to which its approach to speech regulation reflects a conception of autonomy

[92] See *Cohen* (n 41) 21.
[93] ibid 24 (emphasis added).
[94] *Hustler* (n 32) 322.
[95] See *Snyder* (n 39) 460 - 461.
[96] *Rowan v Post Office Dept,* 397 US 728, 738 (1970).
[97] For example, the government may act for the purpose of prohibiting intrusion into the privacy of the home of unwelcome views and ideas that may not be banned in public discourse. See *Cohen* (n 41).
[98] ibid 21.

that is firmly rooted in the speaker. This represents a marked divergence from the European conception of autonomy, which is centred on the dignity of listeners/targets of extreme speech. The theories of Waldron and Heyman are therefore instructive in illuminating the philosophical underpinnings of the conception of dignity in Europe and the ways in which these underpinnings are reflected in relevant doctrine.

In contrast to the United States, dignity features prominently in European human rights instruments. For example, Article 1 of the EU Charter states that '[h]uman dignity is inviolable. It must be respected and protected'.[99] Dignity also appears in the constitutions of several European countries.[100] It also features prominently in the jurisprudence of the ECtHR, which recognises that 'tolerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society'.[101] Of particular relevance here is how dignity functions in the context of speech regulation, in particular, the regulation of extreme speech in public discourse. Like Heyman, the ECtHR conceptualises the contours the right to freedom of speech within a broader framework of rights that is premised on individuals respecting others' equal status as human beings and of members of the community.[102]

Notwithstanding the ECtHR's broad pronouncement that Article 10 applies not only to information or ideas that are favourably received but also to those that 'offend, shock, or disturb',[103] it cautions that it 'may be necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote, or justify hatred based on intolerance provided that any formalities, conditions, restrictions or penalties imposed are proportionate to the legitimate aim pursued'.[104] It further observes that when a particular group is singled out for victimisation and discrimination, laws should protect those characteristics that are essential to a person's identity.[105] For these reasons, the Court permits interference with expression that injures the dignity of particular groups as a legitimate aim under Paragraph 2 of Article 10. By way of example, the Court has accepted that protecting the dignity of

[99] Charter of Fundamental Rights of the European Union [2012] OJ C326/391.
[100] By way of example, Article 1 of the German Constitution states '[h]uman dignity shall be inviable. To respect and protect it shall be the duty of all state authority' (Grundgesetz: Basic Law for the Federal Republic of Germany, 23 May 1949).
[101] *Terentyev v Russia*, App no 10692/09 (ECtHR, 4 February 2019) para 65.
[102] See, e.g., *Gough v United Kingdom*, App no 49327/11 (ECtHR, 23 March 2015).
[103] *Handyside v United Kingdom* (1976) ECHR 5, para 49.
[104] *Erbakan v Turkey*, App no 59405/00 (ECtHR, 6 October 2006) para 56 (internal quotations omitted).
[105] *Vejdeland and others v Sweden,* App no 1813/07 (ECtHR, 9 May 2012).

victims of terrorism is a legitimate aim under Paragraph 2.[106] Waldron's notion of group libel is instructive here as the ECtHR frames issues involving extreme speech as a type of 'defamation' of particular groups in certain contexts. For example, in cases involving Holocaust denial, the Court has held that denying crimes against humanity is one of the most serious forms of 'racial defamation' of Jewish people.[107]

The distinction between speech that causes mere offence and speech that constitutes assaults on dignity, which Waldron and Heyman emphasise in their dignity-based theories, is particularly relevant to the challenge of drawing the appropriate line between freedom and regulation. In certain areas, particularly in the sphere of morality, the ECtHR fails to meaningfully distinguish between these two concepts. This is evidenced by the wide margin of appreciation applied by the Court in cases in which proscribed expression is characterised as offending personal beliefs, particularly religious convictions, as well as its characterisation of 'gratuitously offensive' statements as an infringement of the rights of those who take offence.[108] The application of the margin of appreciation doctrine is examined in more detail in Chapters 3, 4, and 6.

III. Theories of Liberty and Equality and the Practice of Attaching Duties and Responsibilities to the Exercise of Rights in European and American Free Speech Frameworks

The final philosophical divergence between the European and American approach to the regulation of extreme speech that informs this thesis is tied to the practice of attaching duties and responsibilities to the exercise of rights. While it is customary in European constitutional frameworks to conceive of rights as including attendant duties and responsibilities, no such conception of rights exists in the American constitutional tradition. This distinction reflects divergences concerning the philosophical principles that underpin American and European human rights frameworks, in particular, the relevance of liberty and equality.

---

[106] See, e.g., *Leroy v France*, App no 36109/03 (ECtHR, 2 October 2008) (upholding the conviction of a man for publishing a drawing representing the attack on the twin towers of the World Trade Centre with the caption '[w]e all dreamt it…Hamas did it,' determining that through his choice of language, the applicant had expressed moral support for the perpetrators of the September 11, 2001 terrorist attacks on the United States and, in so doing, diminished the dignity of the victims).

[107] See *Garaudy v France,* App no 65831/01 (ECtHR, 24 June 2003); *Pastörs v Germany*, App no 55225/14 (ECtHR, 3 October 2019).

[108] See *Otto-Preminger-Institut v Austria,* App no 13470/87 (ECtHR, 20 September 1994).

A. Free Speech Theories Rooted in Liberty and/or Equality

The precise nature and meaning of liberty and equality in the context of philosophical inquiry, whether they are in harmony or in tension, the extent to which they overlap, if at all, and the roles each should play in democratic orders are the subject of long-standing debates that are beyond the scope of this thesis.[109] Nor does this thesis advocate in favour of one over the other, for example, by positing that a free speech framework rooted in liberty is preferable to one rooted in equality. Rather, my objective here is to provide a high-level sketch of some of the many elements of these theories that assist in understanding the marked differences in the American and European approaches to attaching duties and responsibilities to the exercise of rights, and to examine the ways in which these differences are reflected in the regulation of extreme speech.

Liberty, while a broad term with varying meanings depending on the context, is generally understood as a set of principles based on individuality.[110] One conception of liberty is of being unconstrained by others from doing what one wants.[111] In this context, the primary concern of liberty is less for social benefits than it is for the good of individuals as individuals (rather than as members of a social group). Liberty encourages and values individual choice and raises barriers to governmental or social interference with those goals.[112] Theories of negative liberty recognise a core area in which individuals must be free from state interference if they are to live a truly human life.[113] Thus, liberty is not merely an absence of constraint but a power, that is, the capacity to act in accord with one's own reason and free choice.[114] Most contemporary philosophical debates concerning negative and positive liberty are informed by Isaiah Berlin's *Two Concepts of Liberty.*[115] In broad terms, negative liberty means freedom

[109] As observed by Jan Narveson, '[l]iberty and equality have been discussed over and over, by countless writers, and those discussions have been generally inconclusive'. Jan Narveson, 'Liberty and Equality - A Question of Balance?' in Tibor R Machan (ed), *Liberty and Equality* (Hoover Institution Press 2002) 35.
[110] Frederick Schauer, Free speech: a philosophical inquiry 9 (CUP 1982) 60 - 61.
[111] See James P Sterba, 'Equality is compatible and required by liberty' in Jan Narveson and James P Sterba (eds), *Are Liberty and Equality Compatible* (CUP 2010) 8.
[112] ibid.
[113] Steven J Heyman, 'Positive and Negative Liberty' (1992) 68 Chi-Kent L Rev 8.
[114] ibid 83.
[115] Isaiah Berlin, *Four Essays on Liberty* (1st edn, Oxford Paperbacks 1969) 118. See also Sandra Fredman, *Human Rights Transformed* (OUP 2008); Ronald Dworkin, 'What is Equality? Part 3: The Place of Liberty' (1987) 73(1) Iowa L Rev 1; Charles Taylor, 'What's Wrong With Negative Liberty' in Alan Ryan (ed), *The Idea of Freedom: Essays in Honour of Isaiah Berlin* (OUP 1979) 175 - 93, Susanne Baer, 'Dignity, Liberty, Equality: A Fundamental Rights Triangle of Constitutionalism' (2009) 59(4) U Toronto L J 417.

from - from interference, coercion, or restraint, while positive liberty means freedom *to* - to self-determination, to act or to be as one wills.[116]

Like liberty, equality is a nebulous term to which different meanings attach in different contexts.[117] Its demands are often unclear as is its relation to other political values, including liberty.[118] With that said, unlike liberty, theories of equality are generally concerned with a conceptualisation of rights that is positive in nature.[119] By way of example, [i]f you have a positive right to equality from K and J, then K and J do not have the right to refrain from giving it to you. In one and the same matter, as Thomas Hobbes observes, liberty and obligation are inconsistent: if you are at liberty to do something or not, as you please, then you are not obliged to do one or the other, and if you are obliged to do one or the other, then you are not at liberty to do whichever you like.'[120] In contrast to liberty, equality considerations relate to the social benefits of individuals as members of a social group rather than what is for the good of individuals as individuals. Conceptualised in this way, equality is promoted by way of limiting liberty in the sense that the state moderates the relationships between individuals in order to meet the goal of promoting equality within society. This takes the form of interference with the liberty interests of individual rights-bearers by placing positive obligations on them to respect the rights of others when engaging in the right to freedom of expression.

B. The Relationship Between Theories of Liberty and Equality and Attaching Duties and Responsibilities to the Exercise of Rights in European but not American Free Speech Frameworks

Theories of equality and liberty assist in understanding why the practice of attaching duties and responsibilities to the exercise of rights is commonplace in Europe and expressly rejected in the United States, and the extent to which this is reflected in the protections afforded to the forms of extreme speech discussed in Part II.[121] The United States Constitution is 'a charter

[116] Heyman 'Positive and Negative Liberty' (n 113) 81.
[117] Ronald Dworkin described equality as 'a popular but mysterious ideal'. Ronald Dworkin, *Sovereign Virtue: The Theory and Practice and Equality* (Harvard University Press 2000) 1.
[118] Arthur Ripstein, 'Liberty and Equality' in Arthur Ripstein (ed), *Ronald Dworkin* (CUP 2007) 82.
[119] Narveson (n 109) 44.
[120] ibid.
[121] While conceptions of equality are often linked with notions of dignity, this section focuses on conceptions of liberty and equality in European and American free speech frameworks and how these conceptions manifest in the practice of attaching duties and responsibilities to the exercise of rights in the former and not the latter. This is distinct from the discussion of liberty in the preceding section. For an example of scholarship that discusses the relationship between equality and dignity, see Christopher McCrudden, 'Human Dignity and Judicial Interpretation of Human Rights' (2018) 19 EJIL 4.

of negative liberties; it tells the state to let people alone'.[122] The emphasis on the development of an individual's character, opinion, and belief without government interference reflects the broader notion of negative liberty enshrined in the United States Constitution and the prevailing principle that the government has limited authority to regulate the ways in which rights are exercised.

The emphasis on individual identity in American free speech jurisprudence reflects the extent to which constitutional rights in the United States are understood almost exclusively in terms of the relationship between the individual and the government, specifically, the liberty interest of the individual to be free from government interference.[123] There is no conception of the nature of the relationship between individuals. These principles reflect the enduring ethos of ardent individualism in American culture more broadly, which is hostile to any conception of duties and responsibilities to others in relation to the exercise of rights. Such a conception also departs from one of the most fundamental precepts of American constitutionalism, that 'the freedom secured by the Constitution consists…of the right of the individual not to be injured by the unlawful exercise of governmental power'.[124] For this reason, the USSC has long held that the Bill of Rights is designed to secure individual liberty, that is, it creates a 'substantial measure of sanctuary from unjustified interference *by the State'*.[125] This liberty includes specific rights that allow persons, within a lawful realm, to define and express their identity.[126]

[122] *Bowers v Devito*, 686 F2d 616 (7th Cir 1982). See also Michael Rosenfeld, 'Hate Speech in Constitutional Jurisprudence' in Herz and Molnar (n 14) 247: '[i]n essence, free-speech rights in the United States are conceived as belonging to the individual against the state, and they are enshrined in the First Amendment to the Constitution as a prohibition against government interference, rather than as the imposition of a positive duty on government to guarantee the receipt and transmission of ideas among its citizens'.

[123] This is not to suggest that principles of equality play no part in the American constitutional framework. Indeed, the USSC recognises a principle of equality within the United States Constitution in its interpretation of the Equal Protection Clause of the Fourteenth Amendment, holding that the principle of equality lies at 'the heart of the Fourteenth Amendment'. *Loving v Virginia*, 388 US 1, 11 (1967). The Equal Protection Clause commands that no state shall 'deny to any person within its jurisdiction the equal protection of the laws', which the USSC interprets as 'essentially a direction that all persons similarly situated should be treated alike'. *Cleborne, Texas v Cleburne Living Center*, 473 US 432, 439 (1985). The Court has applied the equality principles underlying the Fourteenth Amendment to invalidate laws banning same-sex marriage (*Obergefell v Hodges*, 576 US 644 (2015)), inter-racial marriage (*Loving v Virginia*) and racial segregation in public schools (*Brown v Bd of Educ of Topeka*, 347 US 483 (1954)). Principles of equality do not figure prominently outside of the Fourteenth Amendment context, however, and even within that context, equality is construed in terms of an obligation on the state to treat similarly situated individuals alike, rather than an obligation on citizens in relation to their treatment of one another.

[124] *Schuette v BAMN*, 572 SCt 1623, 1636 - 1637 (2014).

[125] *Roberts v United States Jaycees*, 468 US 609, 618 (1984) (emphasis added).

[126] *Obergefell* (n 123) 651 - 652 (emphasis added).

The predominant emphasis placed on individual identity in the American constitutional framework may be explained, in part, by the country's historical origins. For example, in distinguishing the First Amendment from the English free speech tradition, the USSC opined that it 'cannot reasonably be taken as approving English practices prevalent at the time of its adoption, but on the contrary the unqualified prohibitions laid down by the framers thereof were intended to give to liberty…the broadest scope that could be countenanced in an orderly society'.[127] The Court further observes that there are no contrary implications in any part of the history in which the First Amendment was framed and adopted, and that 'no purpose in ratifying the Bill of Rights was clearer than that of securing for the people of the United States 'much greater' freedom of expression 'than the people of Great Britain had ever enjoyed'.[128] Additionally, the Court emphasises that the people of the United States 'have ordained in the light of history, that in spite of the probability of excesses and abuses,' that the liberty enshrined in the First Amendment is 'in the long view' essential to democracy.[129]

For all of these reasons, the USSC regards government efforts to moderate the ways in which free speech is exercised in public discourse with suspicion and subjects such regulations to rigorous scrutiny. While the American approach conceptualises constitutional protections as providing a type of shield that affords individuals the space to develop, define, and express their identities absent government interference, particularly when exercising their First Amendment rights in the domain of public discourse, the approach in Europe is markedly different in that European courts moderate the free speech rights of individuals in public discourse in several ways.[130] The limitations on the right to freedom of speech in the European framework reflect a conception of democracy that does not prioritise liberty over other political and social ideals. This is exemplified by the limitations on freedom of expression in paragraph 2 of Article 10, which include duties and responsibilities, that the ECtHR takes into account when assessing the necessity of challenged proscriptions on speech. Thus, while the ECtHR

[127] *Bridges v California*, 314 US 252, 264 - 265 (1941).
[128] ibid. As one federal court observed: '[b]orn from immigrants, our national identity is woven together from a mix of cultures and shaped by countless permutations of geography, race, national origin, religion, wealth, experience, and education. Rather than conform to a single notion of what it means to be an American, we are fiercely individualistic as a people, despite the common threads that bind us. This diversity contributes to our capacity to hold a broad array of opinions on an incalculable number of topics. It is our freedom as Americans, particularly the freedom of speech, which generally allows us to express our views without fear of government sanction.' *Bible Believers v Wayne County, Michigan*, 805 F3d 228, 233 (6th Cir 2015).
[129] *Cantwell v Connecticut*, 310 US 296, 310 (1940).
[130] *See, e.g., Mamère v France*, App no 12697/03) (ECtHR, 7 February 2007) para 26; *Lindon, Otchakovsky-Laurens and July v France,* App nos 21279/02 and 36448/02 (ECtHR, 22 October 2007) para 56.

interprets Article 10 as applying to information or ideas that offend, shock or disturb, it subjects such freedom to 'many restrictions'.[131] And while individuals participating in public discourse are permitted to engage in a 'degree of exaggeration or of even provocation,' in so doing they undertake attendant duties and responsibilities in relation to others.[132] For example, as discussed above, these include refraining from engaging in expression that may be gratuitously offensive to others.[133]

A further limitation on the exercise of the right to freedom of expression in the European framework is that it may not be exercised in an 'irresponsible' or 'excessive' manner.[134] Relevant considerations include whether speech is offensive or humiliating to particular persons or groups, whether it challenges the honour or reputation of others, and whether it was necessary for other individuals to form an opinion about the relevant subject matter.[135] This stems from a conceptualisation of free speech as a 'dialogue and a spirit of compromise, necessarily entailing various concessions on the part of individuals or groups of individuals, which are justified in order to maintain and promote the ideals and values of a democratic society'.[136] This conception of free speech requires the subordination of individual interests, which may reasonably be characterised as liberty interests, to those of the group in certain cases to ensure 'the fair and proper treatment of people from minorities and avoid any abuse of a dominant position'.[137] The latter may be construed as a positive right to equality on behalf of the potential targets of extreme speech. Viewed through the equality lens identified in the preceding section, this approach may be understood as promoting equality by limiting the liberty of those who desire to exercise the right to freedom of expression recklessly, i.e., in ways that undermine the rights of others.

The divergence in the American and European approaches with respect to attaching duties and responsibilities to the exercise of the right to freedom of expression reflect significant differences in conceptions of democracy that may be attributable, at least in part, to the

[131] ibid.
[132] *Maguire* (n 44) para 49.
[133] See, e.g., *Otto-Preminger* (n 108).
[134] See *Vejdeland* (n 105).
[135] See *Palomo Sánchez and others v Spain*, App nos 28955/06, 28957/06, 28959/06 and 28964/06 (ECtHR, 12 September 2011).
[136] *Gough* (n 102).
[137] ibid para 168 (referencing *Chassagnou and Others v France*, App nos 25088/94, 28331/95 and 28443/95 (ECtHR, 29 April 1999) para 112; Leyla *Şahin v Turkey*, App no 44774/98 (ECtHR, 10 November 2005) para 108, and *Bayatyan v Armenia*, App no 23459/03 (ECtHR, 7 July 2011) para 126).

historical circumstances in which this right, and the constitutional frameworks in which it is situated, evolved. This issue touches on fundamental questions concerning the relationship between individuals within a given society and between individuals and the state. These distinctions assist in reaching a deeper understanding of the free speech inquiries carried out in Part II, particularly with respect to the discussion of disinformation in Chapter 5.

Conclusion

This chapter both contextualises and elucidates doctrinal divergences in European and American free speech frameworks by examining the ways in which these divergences reflect significant and enduring differences in the philosophical traditions that underpin these frameworks, for the purpose of enriching and framing the discussions of extreme speech in Part II. With this objective in mind, what observations can be made at this juncture based on the foregoing examination?

Firstly, this chapter highlights that the differing conceptions of the value afforded to public debate reflect deeply rooted disagreements concerning the proper relationship between the government and citizens. Theories of democratic legitimacy view unfettered participation in public discourse as a necessary precondition for such legitimacy. The American approach to speech regulation is premised upon this philosophical conception of legitimacy, as reflected in the USSC's articulation of free speech as the essence of self-government, the rigorous scrutiny it applies to content and viewpoint-based proscriptions on expression, and its acceptance of 'hate speech' and other forms of extreme speech in public discourse. The ECtHR, in contrast, regards the regulation of expression as part of a broader constitutional framework that places equal value on other principles, including tolerance and social cohesion, which justify restrictions on expression to serve such ends. Thus, while citizen participation in public debate is considered an important principle in the European approach to speech regulation, it is one of many others that are balanced against one another, rather than as one that virtually always prevails.

Second, this analysis reveals that differing conceptions of dignity within free speech frameworks reflect distinct conceptions of the way in which autonomy functions in democratic societies. The more rigorous regulation of extreme speech that typifies the European approach results from conceptions of dignity that are firmly rooted in the targets of extreme speech. This helps to explain why the free speech rights of the speakers are balanced against the dignity

rights of the targets and why restrictions on 'hate speech' and terrorist-related expression that falls short of incitement to violence are commonplace at the national level in Europe. In contrast, the American approach is grounded in a conception of autonomy that is rooted in the speaker. Doctrinally, this manifests in a presumption against content and viewpoint-based proscriptions on expression, as well as scepticism regarding government efforts to moderate the types of views permitted in public discourse, regardless of the intentions underlying such moderation.

Third, the firm rejection of placing duties and responsibilities on rights-bearers in the exercise of rights in the American constitutional framework reflects an enduring tradition of negative rights that is underpinned by a commitment to a traditional conception of liberty over conceptions of equality.[138] In contrast, the practice of attaching duties and responsibilities to the exercise of free speech rights that is commonplace at the national and regional levels in Europe reflects an understanding of freedom of expression in which the liberty interests of individual speakers do not prevail over the equality interests of others, including the targets of extreme speech. Rather, the interests of each are balanced against one another.

Finally, what broader insights may be drawn from these observations that inform the examinations undertaken, and arguments made, in Part II? This chapter highlights that the ways that courts interpret the scope of free speech rights are shaped and informed by the broader philosophical traditions that underpin different constitutional systems. Additionally, doctrinal divergences concerning the regulation of extreme speech reflect deeply rooted philosophical disagreements that may be explained, in part, by the socio-historical and cultural climate in which those frameworks developed. For example, the American preoccupation with negative freedom may be understood as a response to its fraught relationship with its former colonial power and the desire for more freedoms following emancipation from British rule. In contrast, the European emphasis on equality, especially at the regional level, may largely be attributed to a post-World War II environment that was preoccupied with resisting totalitarianism and fortifying against future threats to democracy.[139] These doctrinal divergences directly impact the scope of the protections afforded to 'hate speech', terrorist-

[138] This is not to suggest that traditional conceptions of equality are entirely absent from the American constitutional framework but, rather, that liberty is the overwhelmingly dominant principle underpinning the American approach to the regulation of expression.
[139] See Chapter 1, pp 41 - 42.

related expression, and disinformation. The examinations undertaken in this chapter aid in understanding the philosophical roots of these divergences, which grounds the rich and nuanced examinations and observations provided in Chapters 3 through 5, as well as the reflections and recommendations proffered in Chapter 6.

PART II

Introduction to Part II

With the firm doctrinal and philosophical grounding provided in Part I, we turn now to the challenges posed by 'hate speech', terrorist-related expression, and disinformation from state actors in the digital age. Chapters 3 through 5 interrogate government efforts to regulate a particular type of expression against the backdrop of recent developments in Europe and the United States, taking into account how advances in technology continue to transform the contemporary information ecosystem, creating new pathways for communication as well as new challenges and opportunities for regulation. Chapter 6 highlights the themes that emerge from prior chapters and offers some reflections and recommendations.

As emphasised in the Introduction, the decision to examine three types of extreme speech in a single volume was both deliberate and strategic. It reflects the conclusion that the most significant and challenging contemporary free speech questions relate to the online regulation of 'hate speech', terrorist-related speech, and disinformation from state actors, and that examining these types of speech in a single volume provides insights that do not emerge from inquiries directed to a single type of extreme speech. Additionally, engaging with these types of extreme speech within a broader inquiry into the appropriate limits on freedom of expression in the digital age highlights both the unique harms and regulatory challenges flowing from each type of speech while illuminating overlapping and interrelated issues from which broader and more nuanced themes, lessons, and connections emerge. For example, the EU's foray into legislating online terrorist-related expression offers lessons in the transnational free speech implications that arise from compulsory efforts to regulate digital intermediaries and from the use of emerging technologies to moderate content.

Additionally, the apparent costs and benefits of the use of human rights penality in the regulation of 'hate speech' should inform efforts to regulate disinformation, an emerging and complex challenge in the digital age. The examinations undertaken in Part II also expose unsettled and contested elements within and between European and American constitutional frameworks that should inform the discourse surrounding efforts to regulate extreme expression online. The examinations undertaken in Part II are directed to the primary aim of

this thesis, which is to offer insights into the broader discourse concerning the appropriate limits on freedom of democracy in liberal democracies in the digital age.

Chapter 3

Online 'Hate Speech': Applying Longstanding Frameworks Directed to Offline Speech in the Digital Age

---

'Hate persists. It's relentless'.[1]

Introduction

In September 2019, thirty independent UN experts published an open letter urging states and social media platforms to take action to curb the spread of 'hate speech', expressing alarm at the recent increase in hateful expression and incitement to discrimination and hatred against migrants, minority groups and various ethnic groups in several countries.[2] In 2020, the Council of Europe's Anti-Racism Commission raised alarm over 'ultra-nationalistic and xenophobic politics across Europe, hate speech setting the tone in social media, rampant anti-Semitism, and anti-Muslim hatred'.[3] Hate is also on the rise in the United States. In 2019, the Southern Poverty Law Center tracked 940 hate groups across America and identified a fifty-five percent

[1] Heidi Beirich, 'Resolute: The Battle Against Hate Demands Vigilance' (*Southern Poverty Law Center: Intelligence Report,* 10 September 2019) <https://www.splcenter.org/fighting-hate/intelligence-report/2019/resolute-battle-against-hate-demands-vigilance> accessed 20 November 2020.

[2] 'Joint Open letter on concerns about the global increase in hate speech' (23 September 2019) <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25036&LangID=E#:~:text=OHCHR%20%7C%20Joint%20open%20letter%20on,global%20increase%20in%20hate%20speech&text=We%20are%20alarmed%20by%20the,their%20rights%2C%20in%20numerous%20countries.> accessed 2 November 2020.

[3] 'News of the European Commission against Racism and Intolerance' (*Council of Europe*, 27 February 2020) <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/-/ultra-nationalism-antisemitism-anti-muslim-hatred-council-of-europe-s-anti-racism-commission-raises-alarm-over-the-situation-in-europe> accessed 5 November 2020.

increase in white nationalist hate groups from 2017. In 2020, the Brookings Institution described the rise of hate crimes in the contemporary United States as 'an epidemic'.[4]

The distinction between 'hate speech' and legitimate political expression is heavily contested. As Robert Post observes, when the law seeks to combat hate by way of proscribing expression, it is not because the government believes that hate *per se* should be entirely eliminated from society.[5] Rather, it is because the government desires to silence or deter hateful expression directed to identifiable groups, particularly those considered historically vulnerable or marginalised in particular ways in particular contexts.[6] Thus, debates over the regulation[7] of 'hate speech' are not about whether all hateful speech should be proscribed everywhere. Nor do these debates concern whether it is proper for states to regulate expression that incites immediate violence as all liberal democracies, including the United States, proscribe such expression.[8] Rather, the key question is whether hateful expression in public discourse that is unlikely to lead to immediate violence, but that may produce more long-term, subtle, or uncertain harms, should be proscribed.[9]

Additionally, while there is broad consensus that hate is on the rise in Europe and the United States and reducing this social ill is a laudable goal, the extent to which restrictions on freedom of expression, both offline and online, are legitimate or effective means by which to effectuate this result remains heavily contested. Indeed, while the fight against hate 'may be both noble and well-intended, it is also rife with ambivalences, paradoxes, and unsettled concerns that are open to debate'.[10] The UN recently observed that even the characterization of what is 'hateful' is controversial and disputed.[11] Ultimately, these debates are about the appropriate way to combat hate (which every reasonable person agrees is a worthwhile objective) and the extent

[4] Ray Rashawn, 'Addressing hate crimes in Maryland and in America,' (*Brookings Institution*, 7 April 2020) <https://www.brookings.edu/testimonies/addressing-hate-crimes-in-maryland-and-in-america/> accessed 12 November 2020.
[5] See Robert Post, 'Hate Speech' in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy* (OUP 2009) 125-126.
[6] ibid.
[7] Regulation in this context means the application of criminal sanctions as well as civil penalties and other regulatory mechanisms.
[8] Michel Rosenfeld, 'Hate Speech in Constitutional Jurisprudence' in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulations and Responses* (CUP 2012) 247.
[9] ibid.
[10] Thomas Bruhholm and Birgitte Schepelern Johansen (eds), *Hate, Politics, Law: Critical Perspectives on Combating Hate* (OUP 2018) 12.
[11] 'United Nations Strategy and Plan of Action on Hate Speech' (*UN*, May 2019) 2 <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf > accessed 25 June 2021.

to which restrictions on certain rights for the purpose of achieving this objective are legitimate. These debates continue notwithstanding the efforts of countless legal practitioners and scholars to provide answers to the difficult questions posed by efforts to regulate this category of expression. That so many have waded into this discourse, and so many questions remain unresolved, highlights the unique challenges associated with regulating 'hate speech' in human rights frameworks that recognise freedom of expression as an indispensable component of democratic orders.

While the regulation of 'hate speech' in liberal democracies is not a recent phenomenon, rising levels of online 'hate speech' across the globe have exacerbated existing challenges while raising new concerns regarding aggressive regulatory efforts at the national and supranational levels targeting this type of expression. Contemporary regulatory efforts are grounded in frameworks that are burdened by critical unanswered questions, including the normative justifications for regulating this type of expression in the first place and the propriety of using human rights penality in this context. While it is imperative to address these underlying questions in order to gain a full appreciation of the free speech implications of online efforts to regulate 'hate speech', the contemporary discourse often glosses over these issues, instead emphasising the importance of removing 'hate speech' from online platforms. This chapter argues that examining and offering insights into these fundamental questions is a necessary part of debates over the regulation of online 'hate speech', and that ignoring such questions risks applying fundamentally flawed frameworks to broad swaths of online expression.

Relying on the doctrinal and theoretical foundation laid in Part I, the aim of this chapter is to examine and interrogate the ways in which regulations on 'hate speech' are conceptualised and applied in contemporary European and American legal frameworks and the extent to which recent doctrinal and legislative developments in Europe and the United States may inform broader debates regarding the appropriate role of government in regulating online 'hate speech'. Relevant questions include whether adopting a precise definition of 'hate speech' within legal frameworks is preferable to an ad hoc approach, whether criminal sanctions on expression are a proper and effective tool to combat rising levels of societal hate, particularly online, and the free speech implications of granting to states the power to make content and viewpoint-based determinations as to what categories of political expression warrant exclusion from public discourse.

This chapter proceeds as follows. Part I provides an overview of the international framework for the regulation of hateful expression and the influence of international law in this area. Part II examines the ways in which 'hate speech' is defined (or not defined) in European and American legal frameworks and considers the impact on freedom of expression that results from the absence of definitional constraints. Part III analyses the European and American approaches to regulating 'hate speech' with a focus on recent doctrinal and legislative developments, and highlights that notwithstanding marked differences, unpopular political expression that warrants protection at the international level is currently the target of regulation in both Europe and the United States. Ultimately, this chapter concludes that while unpopular political speech deemed 'hateful' by state actors is at risk in both Europe and the United States, the fragmentation and conceptual imprecision of the European approach, which has developed over several generations, coupled with the increasing popularity of criminal sanctions targeting online expression, presents a greater risk to the adequate protection of free speech in public discourse in the digital age.

I. Combatting Hate at the International Level: The International Framework for Regulating 'Hate Speech'

International human rights law reconciles two sets of principles in the context of freedom of expression: the requirement to permit open debate and individual autonomy with the obligation to prevent attacks on vulnerable communities for the purpose of ensuring the equal and non-discriminatory participation of all individuals in public life.[12] This reconciliation occurs absent an agreed upon definition of 'hate speech' in international law.[13] Notwithstanding the lack of clarity surrounding the meaning of 'hate speech', two international human rights instruments direct states to restrict certain types of hateful expression. First, Article 20(2) of the International Covenant on Civil and Political Rights (ICCPR) limits the right to freedom of expression enshrined in Article 19 by instructing that '[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be

[12] 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (*UNGA*, 9 October 2019) UN Doc A/74/486, 4. For a comprehensive examination of the protection of freedom of expression at the international level, see Ivan Hare 'Extreme Speech Under International and Regional Human Rights Standards' in Hare and Weinstein (n 5).

[13] See, e.g., the Council of Europe's acknowledgement that '[t]here is no universally acceptable definition of hate speech' (Steering Committee on Media and Information Society (CDMSI), 'Recommendation CM/rec(2014)6 of the Committee of Ministers to member States on a guide to human rights for Internet users – Explanatory Memorandum' CM(2014)31 (16 April 2014)).

prohibited by law'.[14] The Human Rights Committee has not decisively interpreted these provisions as requiring criminal sanctions, merely stating an obligation to 'provide appropriate sanctions' for violations.[15]

While all States Parties to the ICCPR are required to respect its provisions and implement its framework at the national level, they are permitted reservations and declarations.[16] Unsurprisingly, given its rejection of content and viewpoint-based proscriptions on expression, the United States issued a reservation to Article 20 that it 'does not authorize or require legislation or other action by the United States that would restrict the right of free speech and association protected by the Constitution and laws of the United States.'[17] European countries, including the UK, Denmark, Finland, Belgium, Iceland, and Ireland also made reservations to Article 20.

Second, Article 4 of the International Convention on the Elimination of Racial Discrimination (ICERD) directs States Parties to 'condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination…'[18] Article 4 further requires States to 'declare [as] an offence punishable by law' the dissemination of ideas based on racial superiority or hatred and incitement to racial discrimination'. In so doing, States are instructed to give 'due regard' to the principles set forth in the Universal Declaration of Human Rights and the rights set out in Article 5 of ICERD, which include the right to freedom of opinion and expression.[19]

---

[14] Article 20 also instructs States to prohibit by law '[a]ny propaganda for war' (International Covenant on Civil and Political Rights (adopted 16 December 1996, entered into force 23 March 1976) 999 UNTS 171 (hereinafter 'ICCPR') art 20). Article 19 states that '[e]veryone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice'. Paragraph 3 of Article 19, like Article 10 of the ECHR, places a limitation on the exercise of the right to freedom of expression by attaching duties and responsibilities to the exercise thereof and providing for certain restrictions including respect of the rights or reputations of other, for the protection of national security, public order, public health, or morals.
[15] See 'Prohibiting incitement to discrimination, hostility or violence,' Policy Brief (*ARTICLE 19*, December 2012) <https://www.refworld.org/pdfid/50bf56ee2.pdf> accessed 20 September 2020.
[16] ICCPR (n 14) art 2.
[17] US Senate Committee on Foreign Relations, 'Report on the International Covenant on Civil and Political Rights, S Exec Rep 23, 1 (102d Sess. 1992), reprinted in (1992) ILM 645.
[18] International Convention on the Elimination of All Forms of Racial Discrimination (adopted and opened for signature by General Assembly resolution 2106 (XX) of 21 December 1965, entered into force 4 January 1969) Article 19, 4(a).
[19] ibid.

In ratifying ICERD, the United States asserted that to the extent ICERD seeks to regulate private conduct in a stricter manner than what already exists under American law, it was not obliged to take any such measures.[20] A significant number of European countries also entered reservations or declarations to Article 4, including the UK, Austria, Belgium, France, Ireland, Italy, and Switzerland. There is no international consensus concerning the requirements of Article 4, and its 'due regard' clause leaves open the question of the proper the balance between the right to freedom of expression and the right to freedom from discrimination.[21] While Article 4 is narrower than Article 20 of the ICCPR to the extent that it applies only to racial hatred, it goes substantially beyond the obligations imposed by the latter in its requirement of criminal sanctions rather than mere legal prohibition.[22]

The United States' position as an outlier in its expansive protection of freedom of expression insulates it, with limited but notable exceptions,[23] from censure from international human rights defenders for violating its obligations under international law. In contrast, the aggressive efforts to regulate 'hate speech' at the national and regional levels in Europe have received censure from the UN and civil society organisations for failing to comply with international free speech standards and obligations.[24] These admonitions reflect the broader limitations of international law in protecting human rights, which include the voluntary nature of the international human rights system, the extent to which reservations to human rights instruments serve to dilute their potency, and the weakness of enforcement mechanisms.[25]

Notwithstanding these challenges, international human rights law provides a helpful rubric for evaluating government efforts to regulate 'hate speech'. In particular, the Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (Rabat Plan), while geared toward the ICCPR, provides

[20] US Senate Committee on Foreign Relations, Report on the International Covenant on the Elimination of All Forms of Racial Discrimination, S Exec Rep 103-29 (2d Sess 1994).
[21] 'Prohibiting incitement to discrimination, hostility or violence' (*ARTICLE 19*, December 2012) 13 <https://www.refworld.org/pdfid/50bf56ee2.pdf> accessed 2 May 2020.
[22] See Hare in Hare and Weinstein (n 5) 71 - 72.
[23] One such exception is the current campaign underway at the state and federal levels in the United States to proscribe expression in support of the Boycott, Divestment and Sanctions Movement (BDS Movement). This is discussed in detail infra in Section III(B)(2).
[24] These criticisms are discussed infra in Part III.
[25] See Hare in Hare and Weinstein (n 5) 75.

a useful guide for assessing efforts to regulate this category of extreme speech.[26] Adopted in 2012, the Rabat Plan is the result of a series of consultations convened by the UN High Commissioner for Human Rights on the prohibition of incitement to national, racial, or religious hatred, which explored legislative patterns and judicial practices. One of its primary objectives is to provide for a comprehensive assessment of the state of implementation of the prohibition of incitement in conformity with international human rights law. It recommends that proscriptions on hateful expression focus exclusively on the advocacy of discriminatory hatred that constitutes incitement to hostility, discrimination, or violence - rather than the advocacy of hatred without regard to its tendency to incite action by the audience - and require a likelihood of imminent harm and a reasonable probability of discrimination, hostility, or violence occurring as a direct consequence of such incitement.[27] The UN further advises that only incitement to discrimination, hostility, or violence that meets all six criteria outlined in the Rabat Plan should be criminalised, and that less severe forms of incitement or 'hate speech' warrant civil or administrative law based restrictions or public policy responses.[28]

With respect to state approaches to online hate speech, the UN recommends that 'human rights protections in an offline context must also apply to online speech', that '[t]here should be no special category of online hate speech', and that '[p]enalties on individuals for engaging in unlawful hate speech should not be enhanced merely because the speech occurred online'.[29] Additionally the UN recommends that '[g]overnments should not demand – through legal or extralegal threats – that intermediaries take action that international human rights law would bar States from taking directly'.[30]

[26] 'Addendum to the Annual report of the United Nations High Commissioner for Human Rights' (*UNHCR*, 11 January 2013) UN Doc A/HRC/22/17/Add.4.
[27] The Rabat Plan outlines a six-part threshold test taking into account (1) the social and political context, (2) status of the speaker, (3) intent to incite the audience against a target group, (4) content and form of the speech, (5) extent of its dissemination, and (6) likelihood of harm, including imminence. See 'One-pager on "incitement to hatred"' (*UNHCR*) <https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_threshold_test.pdf> accessed 4 December 2020).
[28] 'Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences' (*UNHCR*, September 2020) <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf> accessed 4 December 2020.
[29] UN 'A/74/486' (n 12) 12, 22.
[30] ibid.

II. Defining 'Hate Speech' in Legal Frameworks in Europe and the United States

While a colloquial understanding of 'hate speech' dominates the contemporary discourse regarding the appropriate limits of free speech in liberal democracies, insufficient attention is paid to how this term is conceptualised and applied in European and American legal frameworks. This section aims to fill this gap by tackling the following questions: to what extent do the European and American courts responsible for adjudicating challenges to governmental interferences with expression regard 'hate speech' as a distinct category of expression? Is it preferable to apply a precise definition of 'hate speech', a vague definition, or no definition at all? What is the legal significance of designating certain types of expression as 'hate speech'? The answers to these questions aid in understanding how 'hate speech' is conceptualised in European and American legal frameworks, to what extent it is distinguished from other categories of speech, and the ways in which such distinctions impact the protection of freedom of expression.

A. Defining (or Choosing Not to Define) 'Hate Speech' in Europe

At present, there is no single definition of 'hate speech' at the EU level.[31] As a result, its precise meaning and scope remains unclear. For example, while the European Commission's fact sheet on the Code of Conduct on countering illegal hate speech online states that 'illegal hate speech' is defined in EU law by Council Framework Decision 2008/913/JHA (Framework Decision)[32] as 'the public incitement to violence or hatred on the basis of certain characteristics, including race, colour, religion, descent and national or ethnic origin', and that its primary objective is to 'ensure that those who engage in illegal hate speech offences…are effectively prosecuted under criminal law by Member States' authorities', the term 'hate speech' does not appear anywhere in the Framework Decision.[33] The European Parliament has expressed that the term 'encompasses any racist, anti-Semitic or homophobic behaviour and/or speech aimed at offending a person because of his/her origin or membership of an ethnic group, nation or

[31] Article 11 of the EU Charter of Fundamental Rights provides the general scope of the right to free speech: 'Everyone has the right to freedom of expression', which 'shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.' (Charter of Fundamental Rights of the European Union [2012] OJ C326/391).

[32] Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law [2008] OJ L328, 55-58.

[33] 'Code of Conduct – Illegal online state speech questions and answers' (*EC Fact Sheet,* June 2016) <https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf> accessed 19 September 2020.

religion, as well as any justification or denying of crimes against humanity, crimes of genocide and war crimes'.[34] The European Commission Against Racism and Intolerance's General Policy Recommendation on Combatting Hate Speech states that, for the purposes of the Recommendation, 'hate speech' is to be understood as:

> The advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of "race", colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status'.[35]

The Council of Europe defines 'hate speech' as constituting 'all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism or ethnocentrism, discrimination and hostility against minorities, migrants, and people of immigrant origin.'[36] The ECtHR has elected not to adopt the Council of Europe's definition of 'hate speech', opting for an 'autonomous conception' of 'hate speech' and analysing alleged interferences on a case by case basis.[37] As discussed below in Section B, this approach has resulted in an underdeveloped and fragmented methodology that poses serious risks to the adequate protection of freedom of expression in Europe.[38]

At the national level in Europe, there is wide variation with respect to how 'hate speech' is conceptualised, with many states failing to officially recognise or codify a definition within

[34] See Directorate-General for Internal Policies of the Union, 'The European legal framework on hate speech, blasphemy and its interaction with freedom of expression' (2015) <https://www.europarl.europa.eu/RegData/etudes/STUD/2015/536460/IPOL_STU(2015)536460_EN.pdf > accessed 18 December 2020.

[35] 'General Policy Recommendation No 15 on Combating Hate Speech' (*ECRI*, 21 March 2016) <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01> accessed 19 September 2020.

[36] Recommendations and Declarations of the Committee of Ministers of the Council of Europe in the field of media and information society' (*Directorate General of Human Rights and Rule of Law*, 2016) 76 - 78 outlining Recommendation No R (97) 20 of the Committee of Ministers to member states on "hate speech" (adopted 30 October 1997).

[37] Francoise Tulkens, 'When to say is to do: freedom of expression and hate speech in the case-law of the European Court of Human Rights', ECtHR European Judicial Training Network: Seminar on Human Rights for European Judicial Trainers (Strasbourg, 9 October 2012) 2.

[38] This discussion builds off the examination of the Court's applications of Article 17 in free speech cases in Chapter 1, pp 41 - 45.

national legislation or guiding principles.[39] Both the UN and civil society organisations have expressed concerns regarding imprecise definitions of 'hate speech' within national legal frameworks. The UN emphasises that the lack of clarity and consensus around the meaning of 'hate speech' enables governments to violate a broad range of protected expression of citizens while using hateful expression to attack political enemies, dissenters, and critics.[40] Amnesty International warns of the pernicious effects of imprecise regulations, stressing that in the absence of a clearly delineated boundary between legal and illegal expression, individuals tend to engage in self-censorship.[41] In so doing, they refrain from lawfully exercising their right to freedom of expression for fear of punishment, which leads to insidious effects on freedom of expression as a whole.[42] Amnesty also emphasises the danger to free speech even in instances in which the laws are not excessively broad or intentionally abused because 'hate speech' is 'a disputed term that invites subjective analysis'.[43]

Of equal concern is the extent to which non-existent or vague definitions of 'hate speech' within legal frameworks lack the requisite precision to enable individuals to regulate their conduct accordingly.[44] In relation to Article 19 of the ICCPR, the UN instructs that laws be formulated with 'sufficient precision to enable an individual to regulate his or her conduct accordingly' and may not confer upon the government unfettered discretion for the restriction of freedom of expression.[45] Similarly, the ECtHR holds that the exceptions to freedom of expression enumerated in Paragraph 2 of Article 10 'presuppose a definition within domestic

[39] See 'Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU Wide Study and National Assessments' (*PRISM*, 2015) <https://ec.europa.eu/migrant-integration/librarydoc/hate-crime-and-hate-speech-in-europe-comprehensive-analysis-of-international-law-principles-eu-wide-study-and-national-assessments> accessed 15 September 2020. The European Commission Preventing, Redressing and Inhibiting Hate Speech in New Media (PRISM) Project is a project co-financed by the European Commission's Fundamental Rights and Citizenship Programme. This report details the findings of a comprehensive study of European and international law principles concerning the prevention and repression of 'hate speech' and a comparative analysis of national legislation and its effectiveness on hate speech across the EU.

[40] 'Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms' (*UNGA*, 9 October 2019) UN Doc A/74/486. The problems associated with a lack of definitional constraints in the context of the regulation extreme speech is one of the main themes of this thesis and is discussed in subsequent chapters.

[41] 'Written contribution the thematic discussion on Racist Hate Speech and Freedom of Opinion and Expression organized by the United Nations Committee on Elimination of Racial Discrimination' (*Amnesty International*, 28 August 2012) <https://www.amnesty.org/download/Documents/24000/ior420022012en.pdf> accessed 28 September 2020, 6.

[42] ibid.

[43] ibid.

[44] See, e.g., 'General Comment No 34: Article 19: Freedom of opinion and expression' (*HRC*, 12 September 2011), UN Doc CCPRC/GC/34, para 25 of which states that, for 'a norm to be characterized as "law", it must be formulated with specific precision to enable an individual to regulate his or her conduct accordingly…A law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution.'

[45] ibid.

law which is sufficiently clear and unambiguous, thus permitting anyone exercising his freedom of expression to act with reasonable certainty as to the consequences in law of his conduct'.[46]

While there may be broad agreement about some instances that clearly constitute 'hate speech',[47] there are invariably cases in which the line is unclear. Thus, it is reasonable to question whether imprecise definitions of 'hate speech', or in the case of the ECtHR, no definition at all, provide adequate safeguards for freedom of expression within legal frameworks that classify 'hate speech' as a distinct category of expression, particularly for those in which regulatory measures developed to target offline speech are applied in ways that capture broad swaths of online expression. These concerns are explored infra in Part III.

B. 'Hate Speech' is Not Regarded as a Distinct Category of Expression in the United States

As discussed in Chapter 1, the American approach to regulating expression is premised on the principle that the government may not proscribe speech because it disapproves of the subject matter or viewpoint expressed.[48] Nor is 'hate speech' one of the categories of speech identified by the USSC as lying outside the full protection of the First Amendment.[49] Accordingly, as Robert Post notes, 'because "hate speech" is not in the United States itself recognized as a distinct constitutional category of speech act, it is never clear what circumstances people have in mind when they speak of the regulation of hate speech in the United States'.[50]

Because no legal significance attaches to the term 'hate speech' in the American constitutional framework, it is unsurprising that this term is rarely used by American courts. For example, the only USSC case in which the term 'hate speech' appears is *R.A.V. v. City of Saint Paul*.[51] In Justice White's concurring opinion, he characterises 'hate speech' as a subset of the fighting words doctrine, arguing that it may be banned without creating the danger of driving

[46] *Sunday Times v United Kingdom*, App no 6538/74 (ECtHR, 18 May 1977).
[47] For example, inciting others to commit violence against those who belong to a particular race based on animus toward that race.
[48] See Chapter 1, pp 25 - 26.
[49] ibid, pp 36 - 40.
[50] Robert Post, 'Interview with Robert Post' in Herz and Molnar (n 8) 12.
[51] 505 US 377 (1992). This case is discussed in detail in Chapter 1. See Chapter 1, pp 36 - 40.

viewpoints from the marketplace of ideas.[52] He further accuses the majority of 'legitimat[ing] hate speech as a form of public discussion'.[53] In Justice Stevens' concurring opinion, he observes that the disputed ordinance 'does not ban all "hate speech"' but, rather only a subcategory of fighting words'.[54] Based on this reasoning, Stevens argued that the ordinance left open and protected a wide range of expression.[55] Notwithstanding the fact that the term 'hate speech' appears nowhere in the majority opinion in *R.A.V.*, a handful of lower courts have characterised the disputed proscription as targeting 'hate speech'. For example, in *Saxe v. State College Area School District*, the Third Circuit interpreted *R.A.V.* as striking down a down 'a municipal hate-speech ordinance prohibiting "fighting words"'.[56]

The 2017 case of *Matal v. Tam* provides an instructive example of the contemporary American approach to government regulations targeting the type of expression that is colloquially understood as 'hate speech', that is, hateful expression directed to minority groups.[57] In this case, the USSC held that the First Amendment protects expression that 'demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground [that] is hateful'. The dispute in *Matal* concerned the United States Patent and Trademark Office's (USPTO) denial of a trademark filed by the lead singer of the rock group 'The Slants', who claimed he chose the moniker in order to 'reclaim' the term 'and drain its denigrating force as a derogatory term for Asian persons'.[58] The trademark was denied under the disparagement clause of the Lanham Act, which prohibited the registration of trademarks that 'disparage ... or bring ... into contemp[t] or disrepute' any 'persons, living or dead', including 'marks that disparage members of a racial or ethnic group'.[59]

---

[52] ibid 401. Justice White does not offer a definition of 'hate speech' and implies rather than declares that the disputed ordinance proscribed this category of expression. The marketplace of ideas is discussed below in Section III(B) and in Chapter 6.
[53] ibid 402.
[54] ibid 436.
[55] ibid.
[56] 240 F3d 200, 207 (3rd Cir 2001). See also *Aguilar v Avis Rent A Car Sys, Inc*, 21 Cal 4th 121, 152 (SC Cal 1999) (stating that *RAV* held that the 'ordinance banning certain hate speech was unconstitutional'.); *Am Freedom Defense Initiative v Washington Metro Area Transit Auth*, 898 FSupp2d 73, 79 - 80 (D Col 2012) (citing *RAV* in support of the court's conclusion that 'while political speech receives the highest form of protection under the First Amendment, hate speech also can receive First Amendment protections' and ultimately holding that posters that an advocacy group sought to place on subway platforms advocating support of Israel and equating all Muslims with 'savages' was protected speech under the First Amendment as both 'political and hate speech'.).
[57] 137 SCt 1744 (2017).
[58] ibid 1754.
[59] Lanham Act (also known as the Trademark Act of 1946) 15 USC para 1052(a). The Lanham Act is a federal statute governing trademark law, including registration, maintenance, and protection of trademarks used in or affecting interstate commerce (paras 1051 - 1127). The Lanham Act provides a federal cause of action for infringement of trademarks registered with the USPTO.

In defending the disparagement clause, the government argued that it served the interest of preventing underrepresented groups from being 'bombarded with demeaning messages in commercial advertising.'[60] Finding this reasoning unpersuasive, the USSC held that 'no matter how the point is phrased, its unmistakable thrust is this: The Government has an interest in preventing speech expressing ideas that offend…that idea strikes at the heart of the First Amendment'.[61] The Court further opined that '[s]peech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express "the thought that we hate"'.[62] For these reasons, the Court struck down the disparagement clause as facially unconstitutional.

It is worth noting that neither the parties nor the Court characterised the disparagement clause as proscribing 'hate speech'. This is both unsurprising and unremarkable given that, as discussed above, there is no relevance to the term 'hate speech' in the American legal lexicon. However, *Matal* is instructive for a few reasons. First, it provides a contemporary example of the USSC's unequivocal and longstanding rejection of proscriptions on expression that target expression based on content or viewpoint alone, regardless of to whom such speech is directed.[63] Second, it highlights the distinction between 'hate speech' and fighting words. The government proffered no argument, and in the context of copyrights no such argument would be tenable, that the disparagement clause was aimed at preventing imminent violence or immediate breaches of the peace. Nor did it target expression in the form of insults expressed in a 'personally provocative fashion'.[64] Rather, it was a viewpoint-based restriction aimed at protecting underrepresented groups from degrading expression in advertising. As such, it clearly did not target 'fighting words'. However, even if it could reasonably be argued that the disparagement clause targeted 'fighting words', it would have suffered the same fate as the contested ordinance in *R.A.V.* because it singled out a particular viewpoint for regulation.

Because there is no legal significance attached to the term 'hate speech' in American law, the concerns relating to the application of vague or imprecise definitions discussed in relation to

[60] *Matal v Tam,* 137 SCt 1744, 1764 (2017).
[61] ibid.
[62] ibid, quoting *United States v Schwimmer,* 279 US 644, 655 (1929) (Holmes J, dissenting).
[63] This point is further developed below in the discussion of *Beauharnais v Illinois.*
[64] See *Cohen v California*, 403 US 15 (1971).

Europe are not present in the United States. However, this is not to suggest that the USSC's jurisprudence in the context of hateful expression is entirely consistent or that all unpopular political speech in the United States is immune from government regulation. Inconsistencies in the USSC's approach to hateful expression directed to minority groups as well as recent legislative attacks on unpopular political speech in American public discourse are examined infra in part III(B).

III. Efforts to Combat Rising Levels of Societal Hate Through Proscriptions on Expression

The UN Special Rapporteur on freedom of opinion and expression has highlighted that the diverse responses to 'hate speech' between and within regions reflect the 'unclear normative environment' surrounding the regulation of this type of expression.[65] This section examines efforts to regulate hateful expression in Europe and the United States within this environment and considers the implications for freedom of expression that result therefrom, particularly in the context of online speech.

A. Increasingly Aggressive Efforts to Regulate 'Hate Speech' in Europe

1. Criminalising 'Hate Speech': A Popular Tool with Questionable Efficacy

To combat rising levels of hatred, the UN advises that States should generally deploy tools other than criminalising and prohibiting hateful expression, such as education, counter-speech and the promotion of pluralism.[66] Amnesty International recommends that criminal measures be used only as a last resort where less restrictive measures have failed, and urges states to avoid exclusive or undue reliance on punitive measures in favour of holistic approaches to combating prejudices and discrimination.[67] Notwithstanding these recommendations, laws criminalising 'hate speech' are commonplace throughout Europe.[68] This trend reflects obligations under regional and international human rights instruments as well as a growing acceptance at the international level that human rights include positive rights. This acceptance

[65] 'Promotion and Protection of the right to freedom of expression and opinion' (*UNSG*, 7 September 2012) UN Doc A/67/3577, 4.
[66] UNGA 'Report of the Special Rapporteur' (n 12) 12.
[67] Amnesty 'Written contribution' (n 41).
[68] European countries that criminalise 'hate speech' include Belgium, Bulgaria, the Czech Republic, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Portugal, Spain, Sweden, and the UK. See PRISM (n 39).

has resulted in, among other things, a shift from a conception of rights as a limitation on state action - which remains the predominate view in the United States - to one in which rights are conceptualised as demanding state action.[69]

The popularity of criminal sanctions targeting 'hate speech' in Europe warrants particular consideration as the application of criminal law in the human rights context raises unique free speech concerns. The obviousness of a relationship between criminal law and human rights should not function to obscure its paradoxical nature.[70] Some have described human rights as both a 'shield' and a 'sword' of criminal law, while others have described the criminal law as both a protection and a threat for human rights.[71] Liora Lazarus's discussion of 'coercive overreach' is particularly instructive. Lazarus argues that an examination of the relationship between human rights and criminal law should account for the inherent ambiguity in that the latter requires both limiting and requiring state coercion.[72] That is, a coercive duty on the state toward the individual at risk of harm and a coercive duty on the state with respect to the perpetrator of said harm. While the former duty is often framed as 'protective', Lazarus cautions that to conceptualise it simply as protective risks 'masking the coercive sting in its tail'.[73]

With respect to the approach of the ECtHR, Lazarus observes that the relevant jurisprudence fails to address significant questions regarding the normative coverage of criminal law in the context of human rights.[74] Similarly, Natasa Mavronicola challenges the assumption built into the ECtHR's doctrinal framework that criminal law is a practical and effective tool of human rights protection.[75] In particular, she cautions that the coercive duties of criminalisation and criminal redress may result in a narrowing or dilution of the stringency of the obligations that the ECHR imposes on Member States and may serve to undermine the practical, effective, and

[69] Liora Lazarus, 'Positive Obligations and Criminal Justice: Duties to Protect or Coerce' in Lucia Zedner and Julian V Roberts (eds), *Principals and Values in Criminal Law and Criminal Justice* (OUP 2012) 136.
[70] Francoise Tulkens, 'The Paradoxical Relationship between Criminal Law and Human Rights (2011) 9 J Intl Crim Just 577.
[71] See e.g., Kresimir Kamber, *Prosecuting Human rights Offences: Rethinking the Sword Function of Human Rights Law* (BRILL 2017); Mattia Pinto, 'Historical Trends of Human Rights Gone Criminal' (2020) 42(4) HRQ 729.
[72] Lazarus (n 69) 137.
[73] ibid 202.
[74] ibid 136.
[75] Natasa Mavronicola, 'Coercive Overreach, Dilution and Diversion: Potential Dangers of Aligning Human Rights Protection with Criminal Law (Enforcement)' in Laurens Lavrysen and Natasa Mavronicola (eds), *Coercive Human Rights: Positive Duties to Mobilise the Criminal Law under the ECHR* (Hart Publishing 2020).

appropriate presumptions that have developed to hold Member States to account.[76] She also highlights that a coercive orientation to positive obligations for the protection of human rights can distract from alternative tools of protection and may result in significant issues being missed or implicitly downplayed by state actors. By way of example, an emphasis on punishing those responsible for specific human rights violations can result in governments sidestepping the more difficult task of addressing the underlying systemic and/or societal factors that contribute to the proliferation of such violations.[77]

Lazarus and Mavronicola are primarily concerned with coercive duties that arise from the assertion of the right to life and the prohibition of torture enshrined in Articles 2 and 3 of the ECHR, respectively. However, their scholarship is instructive here, both to an examination of the potential deleterious impact on freedom of expression that may result from criminal proscriptions on 'hate speech' as well as the potential for an emphasis on human rights penality to obscure the systemic and cultural forces that contribute to the rising levels of societal hatred. At the national level in Europe, criminal proscriptions on 'hate speech' are integrated into existing criminal law frameworks, which vary by country in terms of the characteristics that are protected and the legal tests that are applied.[78] The use of criminal sanctions targeting 'hate speech' at the national level in Europe is closely monitored by human rights defenders, including the UN and civil society organisations. While an exhaustive review of national legislation is beyond the scope of this thesis, the findings of these bodies are useful in sketching the current landscape 'hate speech' regulation at the national level in Europe.

In a 2018 report produced as part of 'Media Against Hate', a Europe-wide campaign initiated by the European Federation of Journalists and a coalition of civil society organisations, ARTICLE 19 shared its findings from a comparative review of government responses to 'hate speech' in Austria, Germany, Hungary, Italy, Poland, and the UK.[79] Government responses

[76] Mavronicola does not take issue with coercive duties generally and her arguments here are limited to the context of the ECtHR's approach to the use of criminal law in this context.
[77] ibid.
[78] For an in-depth examination of legal frameworks and procedures, procedural and reporting mechanisms, and jurisprudence relating to 'hate speech' at the national level in Europe, see PRISM (n 39).
[79] 'Responding to 'Hate Speech': Comparative Overview of Six EU countries (*ARTICLE 19,* 2018) <https://www.article19.org/resources/responding-hate-speech-comparative-overview-six-eu-countries/> accessed 20 September 2020, 16. See also University of Oxford, 'Comparative Hate Speech Law: Annexure (*Oxford Pro Bono Publico*, March 2012) <https://www.law.ox.ac.uk/sites/files/oxlaw/1a._comparative_hate_speech_annex.pdf> accessed 1 September 2020.

were assessed in light of relevant international standards, including Article 20(2) of the ICCPR and the Rabat Plan.[80] Among ARTICLE 19's findings was that while the countries differ in their approach to addressing various types of 'hate speech', they tend to rely on criminalising this type of expression, and the application and interpretation of such provisions is generally inconsistent.[81] With respect to criminal laws and jurisprudence concerning prohibitions on incitement, it identified significant variations across and within the six countries in terms of how incitement is approached and defined in legislation and applied in practice.[82]

Further, ARTICLE 19's examination of the available jurisprudence revealed that the courts in these countries do not apply a specific incitement test, and that it is unclear whether courts and judicial authorities are even aware of the Rabat Plan or its recommendations.[83] This examination also revealed that the legal reasoning of courts in 'hate speech' cases is 'often vague, ad hoc and seemingly lacking in conceptual discipline or rigour'.[84] ARTICLE 19 assumes that the lack of consistency at the national level may be 'partially attributed to the lack of consistent guidance on how to approach 'hate speech' provided by the jurisprudence of the European Court'.[85] With respect to national efforts to tackle online 'hate speech', ARTICLE 19 notes a 2017 inquiry by the UK Home Affairs Parliamentary Select Committee that examined the effectiveness of the existing legislation for 'hate speech' related offences committed online. Among other things, the inquiry found that 'the laws against online hate speech' were vague and out of date or unclear. The inquiry recommended that the UK government 'review the entire legislative framework on hate speech, harassment and extremism to ensure that the laws are up to date'.[86]

ARTICLE 19 cautions that these deficiencies render regulations on 'hate speech' open to political abuse, including against those minority groups that 'hate speech' laws are ostensibly aimed at protecting. It further emphasises that national legal and policy frameworks are 'insufficient to enable effective resolution of inter-communal tensions or poor social cohesion'.[87] Based on its detailed analysis, ARTICLE 19 makes several conclusions, including

[80] ibid (ARTICLE 19) 16.
[81] ibid 6.
[82] ibid.
[83] ibid 18.
[84] ibid 16.
[85] ibid 18.
[86] ibid 22.
[87] ibid 4.

that legislation - in particular criminal law provisions - should be revised in order to comply with international human rights standards, including the high threshold for limitations on expression set out in the Rabat Plan.[88]

The ECtHR's refusal to adopt a definition of 'hate speech' and its unwillingness to apply a precise test for adjudicating Article 10 cases involving hateful expression has resulted in fragmented and methodologically underdeveloped doctrine that fails to adequately protect the targets of national efforts to proscribe 'hate speech'. The 2020 case of *Lilliendahl v. Iceland* provides a recent and instructive example of the potential dangers to the adequate protection of free speech that flow from this approach.[89] In this case, the applicant violated a provision of the Icelandic penal code directed to 'defamation of character and violations of privacy', which stipulated that '[a]nyone who publicly mocks, defames, denigrates or threatens a person or group of persons by comments or expressions of another nature, for example by means of pictures or symbols, for their nationality, colour, race, religion, sexual orientation or gender identity, or disseminates such materials, shall be fined or imprisoned for up to 2 years.'[90] While the term 'hate speech' did not appear in the contested provision, the ECtHR agreed with the Supreme Court of Iceland's conclusion that the provision targeted this category of expression on the basis that 'it was clear from the provision's preparatory works and the international legal instruments by which it was inspired that the concept of "hate speech" was simultaneously a synonym for the sort of expression which the provision penalized and a threshold for the severity which such expression had to reach in order to fall under the provision'.[91]

The applicant in *Lilliendahl* complained, among other things, that his conviction violated his right to freedom of expression under Article 10. The impugned speech concerned a town's municipal council approval of a proposal to strengthen education and counselling in elementary and secondary schools on matters concerning 'those who identify themselves as lesbian, gay, bisexual or transgender'. Below an online article discussing the proposal, the applicant wrote the following comments:

[88] ibid 41. For example, the Rabat Plan recommends that proscriptions on 'hate speech' take into account an intent to incite the audience against a target group.
[89] App no 29297/18 (ECtHR, 11 June 2020).
[90] General Penal Code No. 19/1940 (*Almenn hegningarlög*).
[91] *Lilliendahl* (n 89) para 33.

> We listeners of [Ú.S.] have no interest in any [expletive] explanation of this kynvilla [derogatory word for homosexuality, literally 'sexual deviation'] from [Ó.S.Ó.]. This is disgusting. To indoctrinate children with how kynvillingar [literally 'sexual deviants'] eðla sig ['copulate', primarily used for animals] in bed. [Ó.S.Ó.] can therefore stay at home, rather than intrude upon [Ú.S.]. How disgusting.[92]

In its analysis, the Court provided a general overview of its approach to 'hate speech', opining that in its case law 'hate speech' falls into the two categories. The first category is 'comprised of the gravest forms of "hate speech"', which the Court construes as falling under Article 17 and 'thus entirely excluded from the protection of Article 10'.[93] The second category is 'comprised of "less grave" forms of "hate speech"' that the Court 'has not considered to fall entirely outside the protection of Article 10, but which it has considered permissible for the Contracting State to restrict'.[94] Included in this category is not only speech that explicitly calls for violence or other criminal acts, but also 'attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population', which 'can be sufficient for allowing the authorities to favour combating prejudicial speech within the context of permitted restrictions on freedom of expression'.[95] In cases concerning speech that does not call for violence or other criminal acts, but which the Court considers to constitute 'hate speech', its conclusion is based on an 'assessment of the content of the expression and the manner of its delivery'.[96]

In applying this approach to the applicant's case, the Court held that that the comments fell under the second category and undertook an Article 10 analysis, reasoning that '[a]lthough the comments were highly prejudicial, as discussed further below, it is not immediately clear that they were aimed at inciting violence and hatred or destroying the rights and freedoms protected by the Convention'.[97] It agreed with the Supreme Court's conclusions that the applicant's comments were 'serious, severely hurtful and prejudicial' and promoted 'intolerance and detestation of homosexual persons', that it acted within its margin of appreciation, that its assessment of the nature and severity of the comments was not 'manifestly unreasonable', and that it had adequately balanced the applicant's right to freedom of expression against the more

[92] ibid para 5.
[93] ibid. See also discussion of Article 17 in Chapter 1, pp. 41 - 45.
[94] ibid (*Lilliendahl*) para 33.
[95] ibid para 36.
[96] ibid.
[97] ibid para 26.

general public interest of the rights of gender and sexual minorities. For these reasons, the Court concluded that it was not its place 'to substitute its own assessment of the merits for that of the Supreme Court'.[98] While the Court's approach reflects an application of the principle of subsidiarity, it is arguably overly deferential in its analysis of the Supreme Court's decision regarding the proper balance between the applicant's Article 10 rights and the rights of others.[99]

In addition to providing a recent illustration of the Court's treatment of 'hate speech' in relation to Article 10, *Lilliendahl* provides an example of the dangers to free speech resulting from the Court's underdeveloped methodology in this area. First, it highlights the lack of clarity from the absence of a definition of 'hate speech' and the lack of rigour undertaken by the Court in determining what types of hateful expression constitute 'hate speech'. This is exemplified by the Court's perfunctory analysis of whether the contested provision targeted 'hate speech' and whether the applicant's expression fell within its scope. Moreover, the Court fails to articulate the difference between 'hate speech' and other types of hateful expression and does not identify criteria by which to distinguish between the 'gravest' and 'less grave' forms within this category.[100]

Second, *Lilliendahl* draws attention to the Court's tolerance of the application of criminal sanctions to hateful expression that falls well outside of traditional notions of incitement.[101] The UN considers the following elements to be 'essential' in determining whether expression constitutes incitement to hatred: real and imminent danger of violence resulting from the expression; intent of the speaker to incite discrimination, hostility or violence; and careful consideration by the judiciary of the context in which hatred was expressed, given that international law prohibits some forms of speech for their consequences, and not for their content as such.[102] In determining that the interference with the applicants' Article 10 rights was justified in this case, the Court acknowledged that the key facts before it were

[98] ibid para 47.
[99] For a comprehensive critical analysis of the ECtHR's contemporary application of subsidiarity, see Stijn Smet, 'When Human Rights Clash in "The Age of Subsidiarity": What Role for the Margin of Appreciation?' in Petr Agha (ed), *Human Rights Between Law and Politics: The Margin of Appreciation in Post-National Contexts* (Hart Publishing 2017).
[100] See Jacob Mchangama and Natalie Alkiviadou, '"Hate Speech" Jurisprudence of the ECtHR through a Qualitative and Quantitative Lens' (*ECHR Blog,* 3 November 2020) <https://www.echrblog.com/2020/11/guest-post-hate-speech-jurisprudence-of.html> accessed 2 December 2020.
[101] The Court's approach to adjudicating challenges to the regulation of online terrorist-related expression, discussed in Chapter 4, is similarly misguided and problematic.
[102] 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (*UNGA*, 7 September 2012) UN Doc A/67/357.

distinguishable from prior cases in which it reached the same conclusion. These include that the applicant was a member of the general public commenting on an online news article regarding the decision of a municipal council, not a politician whose comments were published on a prominent platform that would be received by a wide audience[103] and that his comments were not directed to influencing the views of vulnerable groups.[104]

This analysis reflects the Court's failure to meaningfully distinguish between speech that rises to the level of incitement and speech that is merely offensive.[105] Indeed, there was no evidence before the Court in *Lilliendahl* that the applicant intended to incite hatred or violence against homosexuals. Nor was there evidence that any hatred or violence resulted, or was likely to result, from his comments.[106] Rather, he was a member of the public sharing his opinions on an issue of public concern within his community in Iceland. While his speech was undoubtedly offensive, shocking, and disturbing, it would strain credulity to argue that it rose to the level of incitement.[107] These types of distinctions are particularly significant in the context of online speech as the speaker and the forum significantly influence the extent to which expression is disseminated and its impact on the intended audience. Additionally, establishing links between speech and violence may become more difficult to discern in the context of online expression.

---

[103] cf *Féret v Belgium*, App no 15615/07 (ECtHR, 16 July 2009). In *Féret*, the Court found no violation of Article 10 of the Convention in respect of the conviction of the applicant, chairman of the political party 'Front National', for publicly inciting discrimination or hatred. The Court considered it significant that the applicant's racist statements had been made by him in his capacity as a politician during a political campaign, where they were bound to be received by a wide audience and have more impact than if they had been made by a member of the general public.

[104] cf *Vejdeland and others v Sweden,* App no 1813/07 (ECtHR, 9 February 2012). In *Vejdeland*, the Court reasoned that 'it is hard to see the wording of the leaflets simply as starting a debate on an issue concerning a matter of public interest; it appears rather that the applicants wanted to disseminate their views among teenagers, who are vulnerable to different kinds of influence.' ibid para 8.

[105] See Peter Molnar and Jacob Mchangama, 'The Problem with Hate Speech Laws' (2015) 13(1) Rev Faith and Int Affrs 75.

[106] cf *Karastelev v Russia*, App no 16435/10 (ECtHR, 6 October 2020). In this case, also decided in 2020, the Court emphasised that when assessing a specific 'interference' with expression, relevant factors include: the context in which the impugned statements were made and their potential to lead to harmful consequences (such as violent obstruction of lawful activities of public authorities); whether the statements were made against a tense political or social background; whether the statements, fairly construed and seen in their immediate or wider context, could be seen as a direct or indirect call for violence or as a justification of violence (or hatred or intolerance as may be pertinent in other situations); and the manner in which the statements were made, and their capacity – direct or indirect – to lead to such harmful consequences.

[107] For an in-depth and critical analysis of the ECtHR's 'hate speech' jurisprudence, see Natalie Alkiviadou and Jacob Mchangama, 'Hate Speech and the European Court of Human Rights: Whatever Happened to the Right to Offend, Shock, or Disturb?' (2021) 00 HRL Rev 1. In this piece, the authors argue that the Court has adopted an overly-restrictive approach to hate speech that fails to adequately protect controversial political speech while also failing to adopt an adequate test for balancing the impact of 'hate speech' with freedom of expression and simply assuming, rather requiring Member States to demonstrate, that prohibitions on 'hate speech' are the most efficient tool to foster tolerance and social cohesion.

A more comprehensive critique of the European approach to the regulation of 'hate speech' and its implications for the protection of online expression is provided in Chapter 6.

2. The Dangers of Coercive Overreach

The popularity of criminal sanctions targeting online and offline 'hate speech' at the national level, coupled with the absence of a developed and rigorous methodology for adjudicating challenges to government interferences with hateful expression at the ECtHR, raise serious concerns regarding coercive overreach in the digital age. By way of example, at both the national and supranational levels in Europe, insufficient consideration is paid to the inherent ambiguity of human rights law as both limiting and requiring state coercion, with too much emphasis placed on the latter. This leads to an environment in which states are granted too much latitude and discretion in curtailing freedom of expression on the grounds of protecting others from hatred and discrimination.

Additionally, *Lilliendahl* highlights the salience of Mavronicola's observations regarding the ECtHR, in particular its acceptance of criminal law as an appropriate tool to protect human rights and the resulting dilution of the stringent obligations placed on states to respect freedom of expression. Her warning that a coercive orientation to positive obligations for the protection of human rights can distract from alternative tools of protection is equally relevant in the context of 'hate speech' regulation. For example, ARTICLE 19 has observed that while the German government and mainstream parties frequently proclaim that it is necessary to listen to minority groups and to take into account the concerns and fears regarding immigration and multi-culturalism, aside from passing legislation to control public debate, little to nothing seems to have been done by public officials to address these concerns or to meaningfully engage in public discourse.[108] A lack of comprehensive engagement by national governments contributes to a wider failure to address the root causes of hatred.[109]

Finally, it is reasonable to question whether the European preoccupation with prosecuting individual offenders for specific instances of 'hate speech' by way of criminal sanctions may serve to obscure the ways in which such expression reflects deeply ingrained social inequity and discrimination, and that by focusing on and targeting individual offenders by way of a

[108] 'Germany: Responding to "hate speech"' (*ARTICLE 19 Country Report*, 2018) <https://www.article19.org/wp-content/uploads/2018/07/Germany-Responding-to-%E2%80%98hate-speech%E2%80%99-v3-WEB.pdf> accessed 3 November 2020, 9.
[109] ibid.

protective posture a state may claim to be addressing rising levels of hate within its borders without actively engaging in the more difficult work of instituting broader systemic change.[110] Such efforts may also permit a state to ignore the extent to which its own policies, and the behaviour of government officials, contribute to rising levels of social and cultural animus toward the very groups that 'hate speech' laws are aimed at protecting.[111]

3. The Current Landscape of the Regulation of Online 'Hate Speech' in Europe

In the context of online regulation, at the EU level a voluntary and co-regulatory framework exists in the form of the EU Code of Conduct countering hate speech online, which is evaluated through a regular monitoring exercise set up in collaboration with participating parties, including Facebook, Twitter, and YouTube, among others.[112] Recent efforts at the EU level and in certain Member States highlight the dangers to free speech posed by aggressive efforts to regulate 'hate speech' on online platforms. At the time of writing, six countries, including Austria, France, Italy, Spain, and Germany, have special laws on online 'hate speech' or distinctions within 'hate speech' legislation based on whether impugned expression occurs online or offline.[113] For example, certain speech related offences in Spain's penal code carry a higher penalty if committed online.[114] Such provisions have faced criticism for the underlying presumption that all online expression reaches wider audiences than offline

[110] See Nadine Strossen, 'Interview with Nadine Strossen' in Herz and Molnar (n 8) 380 - 381, in which she argues that '[f]aced with a social problem, policy makers repeatedly scapegoat expression as the subject of regulation, but expression is always only a manifestation of a deeper problem. Attacking expression does not deal with the root causes of the problem and does not deal with its concrete manifestations in terms of conduct…If there is discrimination, let's rectify that. If there is violence, let's prevent and remedy it; let's punish the perpetrators.' See also Nadine Strossen, *HATE: Why We Should Resist It with Free Speech, Not Censorship* (OUP 2018).

[111] For example, the UN Committee on the Elimination of Racial Discrimination expressed serious concern at the sharp increase in the number of racist hate crimes in England, Wales, and Northern Ireland in the weeks prior to an following the referendum on the membership of the European Union—also known as Brexit - remarking that the pro-Brexit campaign was 'marked by divisive, anti-immigrant and xenophobic rhetoric, and that many politicians and prominent political figures not only failed to condemn it, but also created and entrenched prejudices, thereby emboldening individuals to carry out acts of intimidation and hate towards ethnic or ethno-religious minority communities and people who are visibly different'. See 'Concluding observations on the twenty-first to twenty-third periodic reports of United Kingdom' (CERD/C/GBR/CO/21-23, 26 August 2016) para 15.

[112] See 'The EU Code of conduct on countering illegal hate speech online: The robust response provided by the European Union', ST 12522/19 COR 1 (*European Commission*, 2 October 2019) <https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play__0.pdf> accessed 20 June 2021.

[113] For a comprehensive overview of national 'hate speech' legislation in EU Member States, see 'Tackling Disinformation and Online Hate Speech: EU and Member State approaches, so far' (*Democracy Reporting International*, December 2020) <https://democracy-reporting.org/wp-content/uploads/2021/01/Tackling-Disinformation-and-Online-Hate-Speech-DRI.pdf> accessed 28 June 2021.

[114] See 'Spain: Speech related offences of the Penal Code' (*ARTICLE 19*, March 2020) <https://www.article19.org/wp-content/uploads/2020/03/Spain-Penal-Code-analysis-March-2020-Final.pdf> accessed 4 August 2020.

expression and for not requiring any determination of the reach of particular content in a given case.[115]

The most controversial development in the regulation of online 'hate speech' in Europe is Germany's Network Enforcement Act (NetzDG),[116] which the Transatlantic Working Group (TWG) describes as 'arguably the most ambitious attempt by a Western state to hold social media platforms responsible for combating online speech deemed illegal under the domestic law'.[117] Among other things, it imposes intermediary liability for social media networks with over two million registered users, which are required to take down illegal content, including 'hate speech'. Any 'manifestly unlawful' content must be removed within 24 hours. Failure to remove illegal content in the required time frame is punishable by fines of up to 50 million euros.[118]

Since its entry into force in 2017, the scope of NetzDG has been widely debated and criticised by human rights defenders as vague and over-inclusive, 'privatizing' online censorship with little transparency or due process, and incentivising digital intermediaries to err on the side of caution rather than freedom of expression.[119] Human Rights Watch highlights two aspects of the law that violate Germany's obligation to respect freedom of expression. First, 'the law places the burden on companies that host third-party content to make difficult determinations of when user speech violates the law, under conditions that encourage suppression of arguably lawful speech.[120] Second, Human Right Watch emphasises that NetzDG fails to provide either judicial oversight or a judicial remedy in the event a decision by an intermediary violates a

[115] ibid.
[116] Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG) (Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)).
[117] Paddy Leersson and Heidi Tworek, 'An Analysis of Germany's NetzDG Law' (*Transatlantic Working Group*, 15 April 2019) 1 <https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf> accessed 28 June 2021. The TWG is a high-level commission that includes government representatives, legislators, corporate and other policy experts from the EU, EU Member States, and the United States. Its purpose is to 'identify and encourage adoption of scalable solutions to reduce hate speech, violent extremism and viral deception online, while protecting freedom of expression and a vibrant, global internet' (see https://www.ivir.nl/twg/).
[118] NetzDG (n 116).
[119] See, e.g., Joelle Fiss and Jacob Mchangama, 'The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship' (*Justitia*, 2019) <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf> accessed 28 June 2021.
[120] 'Germany: Flawed Social Media Law–NetzDG is Wrong Response to Online Abuse' (*Human Rights Watch,* 14 February 2018) <https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law> accessed 28 June 2021.

user's right to speak or access information.[121] In this way, the largest platforms for online expression become 'no accountability' zones, where government pressure to censor evades judicial scrutiny.[122] In July of 2021, Google filed suit in a German administrative court challenging NetzDG, particularly a provision that requires digital intermediaries to share with law enforcement the details of users sharing content suspected to be hateful prior to a determination of any criminal wrongdoing.[123] In commenting on the case, a Google representative stated that '[n]etwork providers such as YouTube are now required to automatically transfer user data en masse and in bulk to law enforcement agencies without any legal order, without knowledge of the user, only based on the suspicion of a criminal offence'.[124]

Daphne Keller, director of the Program on Platform Regulation at Stanford's Cyber Policy Center, cautions that over-removal is increasingly common under laws that incentivise platforms to err on the side of removing content.[125] In so doing, she highlights the dangers of over-regulation resulting from NetzDG, noting that application of the law has already led to 'embarrassing mistakes', including Twitter suspending a German satirical magazine for mocking a politician and Facebook removing a photo of a bikini top 'artfully' draped over a double speed bump sign.[126] Following its enactment, at least 13 countries and the EU[127] have adopted or proposed models of intermediary liability broadly similar to the NetzDG.[128] According to Freedom House's 2019 assessment of freedom on the internet, five of these countries are ranked as being 'not free', five are ranked 'partly free', and three are ranked 'free'.[129] Many of these countries have explicitly referred to the NetzDG as an inspiration or justification for the approach to intermediary liability.[130] The application of such laws and the impact on freedom of expression, warrant rigorous scrutiny. Chapter 4 applies such scrutiny

[121] ibid.
[122] ibid.
[123] Douglas Busvine, 'Google takes legal action over Germany's expanded hate-speech law,' *Reuters* (27 July 2021) <https://www.reuters.com/technology/google-takes-legal-action-over-germanys-expanded-hate-speech-law-2021-07-27/> accessed 21 July 2021.
[124] ibid.
[125] Daphne Keller, 'Internet Platforms: Observations on Speech, Danger, and Money' (2018) Hoover Institution's Aegis Paper Series No 1807 <https://ssrn.com/abstract=3262936> accessed 28 June 2021.
[126] ibid.
[127] This refers to the EU legislation concerning terrorist-related content discussed in Chapter 4.
[128] Fiss and Mchangama (n 119) 17.
[129] Allie Funk and Adrian Shahbaz, 'Freedom on the Net 2019: The Pandemic's Digital Shadow' (*Freedom House*, 2020) <https://freedomhouse.org/report/freedom-net/2020/pandemics-digital-shadow> accessed 25 March 2021.
[130] ibid.

in relation to recent efforts at the regional level in Europe to regulate online terrorist-related content by way of compulsory regulation of digital intermediaries.

B. American Devotion to Combatting Hateful Speech with More Speech, Not Less (Most of the Time)

1. From *Beauharnais* to the Marketplace of Ideas

While the European approach to regulating hateful expression is premised on the principle that less speech in public discourse is an effective way of combatting hate, the American approach is premised on the opposite principle, that is, that the cure for hateful speech is more speech, not less.[131] Indeed, the USSC holds that regardless of how pernicious an opinion or idea may seem, Americans must depend on its correction not from judges or juries but from the competition of other ideas.[132] This principle is often referred to as the marketplace of ideas, a metaphor that was first introduced in 1919 by Justice Oliver Wendell Holmes in his dissenting opinion in *Abrams v. U.S.*, in which he opined that 'the best test of truth is the power of the thought to get itself accepted in the competition of the market…'. [133] In the Court's view, the marketplace of ideas must remain free and open in order to prevent the government from controlling the search for 'political truth'.[134] It warns that the absence of such a marketplace will result in a public that is unable to choose a government that reflects its informed will.[135]

While first introduced in the early twentieth century, the marketplace of ideas did not emerge in majority opinions of the USSC until the 1960s.[136] Prior to this period, the American approach to regulating the type of expression commonly referred to as 'hate speech' was largely in line with that of other liberal democracies. For this reason, while the objectives of

[131] See, e.g., *United States v Associated Press*, 52 FSupp 362, 372 (1943) (in which Judge Learned Hand opined that the First Amendment 'presupposes that right conclusions are more likely to be gathered out of a multitude of tongues, than through any kind of authoritative selection. To many this is, and always will be, folly; but we have staked upon it our all.').
[132] *Gertz v Welch, Inc*, 418 US 323, 339-340 (1974).
[133] 250 US 616, 630, 728 (1919). This is closely linked with the oft-cited principal espoused by Justice Brandeis in 1927, that '[i]f there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence' (*Whitney v California*, 274 US 357, 377 (concurring opinion) (1927)). See also *United States v Alvarez*, 567 US 709, 727 (2012): '[T]he response to the unreasoned is the rational; to the uninformed, the enlightened; to the straight-out lie, the simple truth.'
[134] *Consolidated Edison Co v Pub Serv Comm'n*, 547 US 530, 538 (2011).
[135] *Sorrel v IMS Health, Inc*, 564 US 552, 583 (2011).
[136] See *Red Lion Broadcasting Co v FCC*, 395 US 367, 390 (1969) (holding that '[i]t is the purpose of the First Amendment to preserve an uninhibited marketplace of ideas in which truth will ultimately prevail.').

this thesis do not require a comprehensive review of the evolution of the USSC's extreme speech jurisprudence in this area, a slight detour into the past is appropriate at this juncture in order to contextualise the USSC's contemporary treatment of hateful expression.[137] In 1952, the USSC issued its opinion in *Beauharnais v. People of the State of Illinois.*[138] The dispute in *Beauharnais* concerned a libel statute that criminalised, among other things, the publishing or exhibiting in any public place a lithograph or moving picture that portrayed 'depravity, criminality, unchastity, or lack of virtue of a class of citizens of any race, color, creed or religion and which exposed the same to contempt, derision, or obloquy'.[139] Beauharnais was convicted of violating the statute by exhibiting lithographs that called on public officials 'to halt the further encroachment harassment and invasion of white people, their property, neighborhoods, and persons by the Negro and called for 'self respecting white people in Chicago to unite'.[140]

The Court framed the question presented in *Beauharnais* as whether the 'liberty' protected by the Due Process Clause of the Fourteenth Amendment prevented a State from punishing libellous utterances directed a defined group in addition to an individual.[141] In analysing this question, the Court reasoned that because an utterance directed at an individual could be subject to criminal sanctions, it could not deny the State the power to punish the same utterance directed a defined group - in this case, Blacks - *unless* it could say that the statute was a 'wilful and purposeless restriction unrelated to the peace and well-being of the State'.[142] Finding the State's argument persuasive, the Court reasoned that due to the racial unrest and propaganda in the Chicago area in the preceding decades it could not find that the Illinois legislature was 'without reason in seeking ways to curb false or malicious defamation of racial and religious groups, made in public places and by means calculated to have a powerful *emotional impact* on those to whom it was presented'.[143] The Court further held that it would be 'arrant dogmatism' and outside the scope of its authority to deny that the State legislature could believe that 'a man's job and his educational opportunities and the dignity accorded him may depend as much on the reputation of the racial and religious group to which he willy-nilly belongs, as

[137] For a comprehensive discussion of the evolution of the American approach to the regulation of extreme speech, see Rosenfeld (n 8).
[138] 343 US 250 (1952).
[139] ibid 281.
[140] ibid 252.
[141] The Court reasoned that, '[n]o one would dispute that it was libellous to falsely claim an individual is a rapist, robber, carrier of knives and guns, and user of marijuana.' ibid 257 - 258.
[142] ibid 258.
[143] ibid 262 (emphasis added).

on his own merits'.[144] This justification is akin to Jeremy Waldron's notion of 'group defamation' discussed in Chapter 2, which underpins his theory supporting restrictions on 'hate speech'.[145]

The disputed ordinance in *Beauharnais* targeted the type of expression that is categorised as 'hate speech' in contemporary European legal frameworks, that is, speech that degrades groups based on characteristics such as race and religion. While the USSC has yet to expressly overrule *Beauharnais*, it represents a doctrinal aberration that the Court has, for all intents and purposes, ignored for the last 50 years.[146] Moreover, while the USSC and lower courts have referenced *Beauharnais* as recently as 2021, it is never relied upon for its principle holding that the state may, consistent with the First Amendment, proscribe false speech concerning religious and racial groups.[147] Incredibly, in *R.A.V.*, the USSC referenced *Beauharnais* in its discussion of categorical exclusions (for defamation) to First Amendment protections, entirely ignoring that its central holding is that the state may prohibit 'group libel' in addition to libel of individuals.[148] Notwithstanding the doctrinal uncertainty resulting from the fact that *Beauharnais* remains good law, the prevailing approach of the contemporary USSC is that if speech does not fall into one of the limited categorical exclusions discussed in Chapter 1, the government may not proscribe expression based on content or viewpoint, including the imposition of criminal sanctions.[149]

---

[144] ibid 261. The Court also held that the State legislature was reasonable in concluding that 'the wilful purveyors of falsehood concerning racial and religious groups promote strife and tend powerfully to obstruct the manifold adjustments required for free, ordered life in a metropolitan, polyglot community.' ibid 263.

[145] See Chapter 2, pp 68 - 69.

[146] See, e.g., *Nuxoll v Indian Prairie Sch Dist # 204*, 523 F3d 668, 672 (7th Cir 2008) (acknowledging that while *Beauharnais* 'has never been overruled, no one thinks the First Amendment would today be interpreted to allow group defamation to be prohibited'); *Dworkin v Hustler Magazine, Inc*, 867 F2d 1188, 1200 (9th Cir 1989) (opining that while *Beauharnais* may be regarded as providing 'some support' for group libel claims, subsequent USSC cases, including *New York Times v Sullivan*, have so weakened its holding that 'the permissibility of group libel claims is highly questionable at best').

[147] See, e.g., *Same Condition, LLC v Codal, Inc*, 2021 IL App 1st 201187, 11 (2021) (referencing *Beauharnais* for the principle that '[a]s a posttrial remedy to prevent the repetition of statements judicially adjudicated to be defamatory, this necessarily means that the statements have already been found to be defamatory and thus are not constitutionally protected').

[148] *Beauharnais* (n 138) 382. As discussed in detail in Chapter 1, in *RAV* the USSC struck down a city's 'Bias-Motivated' Crime Ordinance that prohibited the display of a symbol that aroused 'anger, alarm or resentment in others on the basis of race, color, creed, religion or gender' notwithstanding the government's argument that only a content-based measure would communicate to minority groups that the 'group hatred' aspect of such speech was not condoned by the majority. See Chapter 1, pp 36 - 40.

[149] These include fighting words, true threats, obscenity, and defamation. See Chapter 1, p 36.

The marketplace of ideas aids in understanding why the American approach to speech regulation is generally permissive of 'hate speech' within public discourse and why this type of expression is considered less harmful than government regulation.[150] While the judicial branch remains steadfastly committed to this cornerstone of America's free speech paradigm, recent legislative efforts at the federal and state levels to regulate expressive conduct in support of the BDS Movement and to ban Critical Race Theory (CRT) in public education expose cracks in the foundation of America's free speech paradigm that imperil the protection of politically unpopular expression that the government classifies as hateful.

2. When the Marketplace of Ideas Collides with Unpopular Political Speech

The distinction between 'hate speech' and legitimate political expression is relevant within legal frameworks that recognise the former as a distinct category because the way in which speech is classified distinguishes speech that may be expressed in public discourse and speech that the government may be justifiably exclude therefrom. This section examines recent efforts by legislators at the federal and state levels in the United States to restrict expressive conduct in support of the BDS Movement and to ban CRT from the public education curriculum. This examination highlights both the profound impact that government classifications of unpopular political speech have on the perceived legitimacy of such expression and, by extension, political movements, and the ways in which government actors, even in a jurisdiction like the United States that affords expansive protections to hateful expression, endeavour to silence unpopular political speech under the guise of protecting groups from hatred and discrimination.

While the purpose of the BDS Movement is disputed within American public discourse, its stated aim is to encourage civil society organizations and individuals to use boycotts and divestment initiatives against Israel with the objective of exerting non-violent pressure on the Israeli government to comply with international law and end its activities in the Occupied Palestinian Territory (OPT).[151] The human rights abuses and violations of international law resulting from Israel's actions in the OPT are well documented by a number of international

[150] The extent to which the marketplace of ideas remains fit for purpose in the digital age is examined in Chapter 6.
[151] See the official website of the BDS Movement at <https://bdsmovement.net/what-is-bds> accessed 3 June 2020.

bodies, including the International Criminal Court, the UN Security Council, the UN General Assembly, and the Human Rights Council.[152]

The EU takes the position that actions in support of the BDS Movement are protected speech under the EU Charter.[153] The vice-chair of the European parliament's delegation for relations with Palestine stated in 2016 that, while she did not personally support the BDS movement, '[t]here is an evident wish to silence BDS advocates in order to protect the illegal policies of annexation and dispossession of the Netanyahu government. Criminalising and repressing the legitimate expression of free speech cannot be accepted in our societies'.[154] The vice-chair also rejected 'the unceasing attempts to amalgamate this Palestinian-led movement with antisemtisim'.[155] The discourse surrounding the BDS Movement in the United States is markedly different. In order to appreciate the underlying reasons for this divergence, it is important to briefly contextualise the unique relationship between the United States and Israel. Grounded in the commonality of political and strategic interests, each country occupies a special position in the other's domestic and foreign policy programmes.[156] As a key ally and strategic partner of the United States, Israel is the largest cumulative recipient of American foreign assistance since World War II.[157] As of 2020, the United States had provided Israel $142.3 billion dollars in bilateral assistance and missile defence funding.[158]

---

[152] See, e.g., Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion, ICJ Reports 2004, 136; UN Security Council, Resolution 465 (1980) (S/RES/465, 1 March 1980); UNSC, Resolution 2334 (S/RES/2234, 23 December 2016); 'Resolution on Israeli settlements in the Occupied Palestinian Territory, including East Jerusalem, and the occupied Syrian Golan' (*UNGA,* 23 December 2016) UN Doc A/RES/71/97. In a 2020 report, the UN High Commissioner for Human Rights regarding the Israeli Settlements in the OPT concluded that the establishment and expansion of settlements in the OPT amounts to 'the transfer by Israel of its population …which is prohibited under international humanitarian law' and constitutes 'a war crime that may engage the individual criminal responsibility of those involved.' 'Report on the Israeli settlements in the Occupied Palestinian Territory, including East Jerusalem, and in the occupied Syrian Golan' (30 January 2020) UN Doc A/HRC/43/67.

[153] See 'Answer given by Vice-President Mogherini on Behalf of the Commission (Question Reference E-005122/2016)' (*Parliamentary Questions*, 15 September 2016): 'The EU stands firm in protecting freedom of expression and freedom of association in line with the Charter of Fundamental Rights of the European Union, which is applicable on EU Member States' territory, including with regard to BDS actions carried out on this territory'.

[154] Arthur Nelson, 'European parties urged to agree Israel boycott tactics are antisemitic' *The Guardian* (24 October 2018) <https://www.theguardian.com/world/2018/oct/24/european-parties-urged-agree-israel-boycott-bds-antisemitic-mep> accessed 14 April 2019.

[155] ibid.

[156] For an in-depth examination of the development and nature of the special relationship between the United States and Israel, see Elizabeth Stevens, *US policy towards Israel: the role of political culture in defining the 'special relationship'* (Sussex Academic Press 2006); Abraham Ben-Zvi, *The United States and Israel: The Limits of the Special Relationship* (Columbia University Press 1993); Douglas Little, 'The Making of a Special Relationship: The United States and Israel, 1957-68' (1993) 25 Int J Middle East Stud 563.

[157] Congressional Research Service, 'Report on US Foreign Aid to Israel' (RL33222, 16 November 2020) <https://fas.org/sgp/crs/mideast/RL33222.pdf > accessed 26 January 2020.

[158] ibid.

In the United States, efforts to silence criticism of Israel's actions in the OPT take the form of anti-BDS laws and executive orders in individual states and proposed anti-BDS legislation in Congress. In 2019, more than 250 million Americans, approximately 78 percent of the population, lived in states with anti-BDS laws or policies.[159] The justification for such restrictions is often couched in terms of protecting Israel and the characterisation of the BDS Movement as inherently anti-Semitic. At the time of writing, twenty-seven states have adopted laws or policies that penalise businesses, organisations, and/or individuals that engage in or call for boycotts against Israel and/or Israeli settlements. Often, such laws require that state contractors, including teachers, lawyers, and newspapers, officially certify that they are not participating in boycotts of Israel or companies that do business in Israel.[160]

At the federal level, there are ongoing efforts to enact legislation aimed at the BDS Movement. In 2019, Republican Senator Marco Rubio introduced the 'Strengthening America's Security in the Middle East Act of 2019' (2019 Act), which permits and encourages states to pass anti-BDS legislation.[161] Title IV of the 2019 Act, entitled 'Combatting BDS Act 2019', would allow states to prohibit the investment of state funds in businesses engaged in boycotting Israel and to restrict contracting with any entity that takes part in activities in support of the BDS Movement. Such activities are defined as any action 'intended to penalize, inflict economic harm on, or otherwise limit commercial relations with Israel...for purposes of coercing political action by, or imposing policy positions on, the Government of Israel'.[162] In 2020, Republican Representative Lee Zeldin introduced the Israel Anti-Boycott Act, which would add an amendment to the Export Administration Act of 1979 prohibiting boycotts against allies, including Israel, and block requests request for boycotts by international governmental organisations.[163] In a statement announcing the bill's introduction, Representative Zeldin stated '[w]e have witnessed the rise of anti-Semitism and anti-Israel hate throughout the

[159] 'US: States Use Anti-Boycott Laws to Punish Responsible Businesses' (*Human Rights Watch*, 23 April 2019) <https://www.hrw.org/news/2019/04/23/us-states-use-anti-boycott-laws-punish-responsible-businesses> accessed 30 November 2020.
[160] ibid.
[161] US Congress, 'Strengthening America's Security in the Middle East Act of 2019 (ROS19020)' (116th Cong 1st sess, 2019-2020).
[162] ibid s 402(b).
[163] US Congress, 'Israel Anti-Boycott Act (HR 5595)' (116th Cong, 2019-2020).

world…under the guise of the BDS movement, and whether this bigotry is brazen or it's blatant anti-Semitism deceptively called 'legitimate' we must crush it wherever it exists'.[164]

Efforts to silence pro-BDS expression in the United States have drawn sharp rebukes from both domestic and international human rights defenders. The American Civil Liberties Union and other domestic human rights organisations have brought legal challenges to anti-BDS laws, correctly arguing that they discriminate against disfavoured political expression in violation of the First Amendment and contravene the firmly entrenched principle that participating in politically motivated boycotts is constitutionally protected speech.[165] At the time of writing, federal courts in Texas, Arizona, Georgia, and Kansas have blocked anti-BDS laws on First Amendment grounds.[166] In granting a preliminary injunction, the Kansas court reiterated that political boycotts are 'inherently expressive' conduct protected by the First Amendment.[167] It further emphasised that the law's fundamental goal of undermining the message of those participating in a boycott of Israel was either viewpoint-based discrimination against the opinion that Israel mistreats Palestinians or content-based discrimination on the topic of Israel, both of which are impermissible grounds for proscribing expression under the First Amendment.[168] The Georgia court noted that 'it is easy enough to associate plaintiff's conduct with the message that the boycotters believe Israel should improve its treatment of Palestinians'.[169] It subjected the anti-BDS law to strict rather than intermediate scrutiny because it was a content-based law that targeted speech based on its 'communicative content'.[170]

---

[164] Congressman Lee Zeldin, 'Rep. Zeldin Leads 60 Members in Introducing Israel Anti-Boycott Act (Press Release)' (14 January 2020) <https://zeldin.house.gov/media-center/press-releases/rep-zeldin-leads-60-members-introducing-israel-anti-boycott-act> accessed 29 June 2021.

[165] The recognition of boycotts as legitimate political expression is well established in the United States. See *NAACP v Claiborne Hardware Co*, 458 US 866, 908 (1982) (holding that political boycotts, in which persons 'band[] together and express[] their dissatisfaction with a social structure that ha[s] denied them right to equal treatment and respect'– are protected speech under the First Amendment and that through the exercise of these First Amendment rights, individuals use 'speech, assembly, and petition – rather than [] riot or revolution to…[seek] to change a social order…').

[166] See, e.g., *Koontz v Watson*, 283 FSupp3d 1007 (D Kan 2018); *Jordahl v Brnovich*, 336 FSupp3d 1016 (D Ariz 2018); *Martin v Wrigley*, 2021 WL 2068261 (D Missou 2021); *Anawi v Pflugerville Indep Sch Dis*, 373 FSupp3d 717 (WD Tex 2019). See also overview of anti-BDS legislation at the State level in Brian Hauss, 'Laws Suppressing Boycotts of Israel Don't Prevent Discrimination – They Violate Civil Liberties' (*ACLU*, 22 February 2019) <https://www.aclu.org/blog/free-speech/laws-suppressing-boycotts-israel-dont-prevent-discrimination-they-violate-civil> accessed 28 June 2021.

[167] *Koontz* (ibid) 1023. For an overview of the protection of 'inherently expressive' conduct under the First Amendment, see Chapter 1, pp 30 - 33.

[168] See *Koontz* (n 166) 1022.

[169] *Martin* (n 166) 6.

[170] ibid 7.

The UN Special Rapporteur on the promotion and protection of the right to freedom of expression and opinion published a letter shortly after the introduction of the 2019 Act in which he expressed serious concerns with respect to America's obligations to protect and promote freedom of expression and that the 2019 Act, if enacted, would 'codify a worrying trend of supressing political expression in the country', which includes anti-BDS legislation at the state level.[171] Specifically, the Special Rapporteur raised observations and concerns regarding the compatibility of the 2019 Act and state anti-BDS legislation with Article 19(3) of the ICCPR and urged the federal and state governments to reconsider anti-BDS laws, to not adopt Title IV of the 2019 Act, and to ensure the compliance of any legislation with the United States' obligations under international human rights law.[172] Among these concerns are that government efforts at the federal and state levels 'appear clearly aimed at combatting political expression advocating boycotting, which has long been understood as a legitimate form of expression, protected under Article 19'. The Special Rapporteur expressed particular concern regarding the apparent aim of the 2019 Act to restrict political expression by enacting economic policies that present individuals with an untenable choice: 'Stay silent and avoid BDS advocacy or expect serious negative economic impact…such penalties on individuals or groups who boycott for political reasons will chill political expression and protest, potentially well beyond the strict definition of BDS'.[173]

The ongoing campaign to quash support for the BDS Movement in the United States contravenes firmly entrenched domestic and international free speech standards. The flagrant disregard of core American free speech principles by legislators is particularly troubling. Even if one agrees with the heavily contested premise that the BDS Movement is inherently anti-Semitic, supporters of such measures cannot escape the fact that government proscriptions that expressly target speech based on the views expressed, regardless of how those views are characterised, violate the First Amendment. Indeed, as discussed in Chapter 1, while the USSC acknowledges that political speech is powerful, it firmly rejects the proposition that the government may react to such speech by punishing the speaker for the message conveyed. Moreover, as recently reaffirmed in *Matal*, the American approach to free speech is premised

[171] David Kaye, 'Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (14 February 2019) OL USA 2/2019 <https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24338> accessed 30 November 2020.
[172] ibid 4.
[173] ibid.

on the principle that the First Amendment protects the freedom to express 'the thought that we hate', including hateful expression 'that demeans on the basis of race, ethnicity, gender, religion, age, or disability.'[174]

Speech on matters of public concern is protected so as to ensure the free flow of information into the marketplace of ideas. There is no exception to this general principle for speech that government actors deem hateful. Thus, while government efforts to stifle criticism of Israel's activities in the OPT are packaged as efforts to combat rising levels of anti-Semitism, regardless of the underlying motive, such efforts constitute a serious threat to the protection of political speech in the United States. Not only do such efforts represent a marked retreat from the American commitment to free and open debate on matters of public concern, they serve to delegitimise criticism of Israel's human rights record in the OPT in American public discourse.

Recent events suggest that regulatory efforts to supress unpopular political expression in the United States are not limited to activities in support of the BDS Movement and that the robust protection for unpopular political speech in the United States is at risk. At the time of writing, several State legislatures have introduced bills directed to banning CRT from being taught in classrooms at public education institutions.[175] CRT is a framework developed by American law professors in the 1970s that interrogates the role of race and racism in American society and critiques the ways in which the social construction of race and institutionalised racism perpetuate historical inequalities, recognising the intersectionality of identity.[176] Recent efforts to ban CRT build off of former President Trump's September 2020 Executive Order that excluded from federal contracts diversity and inclusion training 'divisive' content, including

[174] *Matal* (n 60) 1764 (internal quotations and citations omitted).
[175] See Emerson Sykes and Sarah Hinger, 'State Lawmakers are Trying to Bank Talk About Race in Schools' (*ACLU,* 14 May 2021) <https://www.aclu.org/news/free-speech/state-lawmakers-are-trying-to-ban-talk-about-race-in-schools/> accessed 29 June 2021.
[176] See Janel George, 'A Lesson in Critical Race Theory' (*American Bar Association*, 12 January 2021) <https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/civil-rights-reimagining-policing/a-lesson-on-critical-race-theory/> accessed 20 June 2021.

CRT.[177] These laws do not attempt to define CRT but, instead, rely on vague and inaccurate generalities.[178]

In June of 2021, Florida's State Board of Education voted unanimously to adopt a new rule banning lessons that employ CRT or the *New York Times*' 1619 Project[179] in public schools.[180] The new rule states, among other things, that '[e]xamples of theories that distort historical events and are inconsistent with State Board approved standards include the denial or minimization of the Holocaust, and the teaching of Critical Race Theory, meaning the theory that racism is not merely the product of prejudice, but that racism is embedded in American society and its legal systems in order to uphold the supremacy of white persons'.[181] The rule further dictates that '[i]nstruction may not utilize material from the 1619 Project'.[182] Oklahoma's anti-CRT bill bans employees from, among other things, 'mak[ing] part of a course' the concept that 'meritocracy or traits such as a hard work ethic are racist or sexist or were created by members of a particular race to oppress members of another race'.[183] It also bans 'any form of mandatory gender or sexual diversity training of counselling' for any enrolled student of an institution of higher education. Idaho's law bans CRT instruction in all public schools in the state based on the argument that it 'exacerbate[s] and inflame[s] divisions on the basis of sex, race, ethnicity, color, national origin, or other criteria in ways contrary to the unity of the nation and the well-being of the state of Idaho and its citizens'.[184]

---

[177] President Biden rescinded this Executive Order in January of 2021. See 'Executive Order 13985: Advancing Racial Equity and Support for Underserved Communities Through the Federal Government' (*White House Briefing Room*, 20 January 2021) <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/> accessed 20 June 2021.

[178] For example, the Idaho law prohibits 'certain tenets' such as that 'any sex, race, ethnicity, religion, color, or national origin is inherently superior or inferior' that are 'often found in "critical race theory"', Legislature of the State of Idaho, House Bill No 377 (2021) <https://legislature.idaho.gov/wp-content/uploads/sessioninfo/2021/legislation/H0377.pdf> accessed 20 June 2021.

[179] The 1619 Project is an initiative from the *New York Times Magazine* that began in 2019 on the 400th anniversary of the beginning on slavery in the United States, which 'aims to reframe the country's history by placing the consequences of slavery and the contributions of black Americans at the very center of [the] national narrative' (*New York Times Magazine,* The 1619 Project) <https://www.nytimes.com/interactive/2019/08/14/magazine/1619-america-slavery.html> accessed 29 June 2021).

[180] Jacob Oliva, 'Rule 6A-1.094124, Florida Administrative Code (F.A.C.), Required Instruction Reporting' (*Florida Department of Education,* 14 June 2021) <https://info.fldoe.org/docushare/dsweb/Get/Document-8724/dps-2019-165.pdf> accessed 29 June 2021.

[181] ibid.

[182] ibid.

[183] Kevin West and others, 'Enrolled House Bill No 1775' <http://webserver1.lsb.state.ok.us/cf_pdf/2021-22%20ENR/hB/HB1775%20ENR.PDF> accessed 20 June 2021.

[184] House Bill no 377 (n 178).

Efforts to suppress CRT in public schools in the United States raise concerns regarding the broader implications of efforts to silence unpopular political expression under the guise of protecting particular groups from hatred and intolerance, and demonstrate that such efforts are not limited to the BDS Movement. For example, in remarks made in support of the Florida ban, Governor De Santis tweeted that 'Critical Race Theory teaches kids to hate our country and to hate each other. It is state-sanctioned racism and has no place in Florida schools'.[185] In support of Louisiana's anti-CRT bill, Republican State Congressman Ray Garofalo stated that teaching CRT 'furthers racism and fuels hate'.[186]

Academic freedom, like the right to participate in political boycotts, has long enjoyed First Amendment protection. For example, in 1967, the USSC struck down a New York law that required teachers employed by public institutions of higher education to answer, under oath, the following question: 'Have you ever advised or taught or were you ever a member of any society or group of persons which taught or advocated the doctrine that the Government of the United States or of any political subdivisions thereof should be overthrown or overturned by force, violence or any unlawful means?'[187] While acknowledging the legitimacy of the state's interest in protecting the education system from 'subversion' as 'legitimate and substantial', the Court held that such interest 'could not be pursued by means that broadly stifle fundamental personal liberties when the end can be more narrowly achieved'.[188] It went on to emphasise the importance of academic freedom to the First Amendment:

> Our Nation is deeply committed to safeguarding academic freedom, which is of transcendent value to all of us and not merely to the teachers concerned. That freedom is therefore a special concern of the First Amendment, which does not tolerate laws that cast a pall of orthodoxy over the classroom. The vigilant protection of constitutional freedoms is nowhere more vital than in the community of American schools. The classroom is peculiarly the 'marketplace of ideas.' The Nation's future depends upon leaders trained through wide exposure to that robust exchange of ideas which discovers truth out of a multitude of tongues, (rather) than through any kind of authoritative selection.[189]

[185] 10 June 2021.
[186] Melinda Deslatte, 'Bill targeting critical race theory divides La. Lawmakers' *AP News* (27 April 2021) <https://apnews.com/article/education-race-and-ethnicity-government-and-politics-5c4a42900a3d17756d64b6d435f9f41f> accessed 29 June 2021.
[187] *Keyishian v Bd of Regents*, 385 US 589 (1967).
[188] ibid 602, quoting *Shelton v Tucker*, 364 US 479, 488 (1960).
[189] ibid 603 (internal quotations omitted).

Anti-CRT laws are predicated on the principle that educating the public about racism and injustice are, in and of themselves, acts of injustice and racism directed to those who benefit from historical inequities. Like ant-BDS legislation, anti-CRT laws are packaged as efforts to protect groups from hatred and to promote social cohesion that provide a thinly veiled cloak of legitimacy that obscures the true aim of such regulations, which is to wrest control of public debate for the purpose of stifling unpopular speech that challenges dominant political narratives.[190] Human rights defenders, regardless of their personal views concerning the BDS Movement or CRT, should be deeply concerned by recent efforts to silence and delegitimise constitutionally protected expression under the guise of protecting certain groups from hatred, and by the ways in which such efforts may broadly impact the protection of unpopular political speech in the United States.

Conclusion

This chapter highlights problems in contemporary European and American approaches to the regulation of hateful expression, revealing the extent to which fundamental questions concerning the regulation of 'hate speech' in public discourse remain normatively and doctrinally contested both within and across legal frameworks, and that important and longstanding questions remain unanswered. These questions include where to draw the line between offence and incitement and the normative justifications for using the coercive force of the state to silence or punish particular viewpoints.[191] The dangers to free speech posed by the lack of clarity in such frameworks are exacerbated by increasingly aggressive efforts in Europe to regulate 'hate speech' on online platforms. The value in identifying and examining these problems extends beyond the regulation of 'hate speech' to the other forms of extreme speech examined in this thesis.

While international free speech standards provide useful guidance in this area, many European countries, the EU, and the Council of Europe, as well as the United States have demonstrated a general disregard of the international framework in this context. This is particularly concerning given the UN's observation that governments often exploit ambiguities in

[190] That these efforts are exclusively undertaken by members of the Republican Party is unsurprising given that it is one of the most illiberal parties in any contemporary liberal democracy on the planet. Recent research regarding the Republican Party's contemporary shift to illiberalism and the dangers to American democracy flowing therefrom are discussed in Chapter 5.
[191] The dangers of applying human rights penality in this context are explored further in Chapter 6.

international human rights law in ways that threaten legitimate expression, including political dissent and criticism. This highlights a broader theme in this thesis, which is the inherent danger to freedom of expression that arises from expansive government authority to classify and regulate speech in public discourse and the ways in which these dangers are exacerbated in the context of online speech. Such authority provides states with enormous power to silence and delegitimise dissenting or political unpopular viewpoints - for example, by classifying the expression of certain views as hateful - while enhancing and promoting others.

Many may be comfortable with the government classifying the expression of Neo-Nazis as 'hate speech' and unworthy of inclusion in public discourse. However, there are many more difficult cases in which the line is not so clear. Attacks on the BDS Movement and CRT in the United States highlight the extent to which unpopular political expression is at risk of regulation even within legal frameworks that provide robust protections to extreme speech. Thus, while it may be tempting to conclude that Europe has gone too far in one direction by aggressively regulating 'hate speech' in the absence of firm normative and doctrinal frameworks, and the United States has gone too far in the other direction by providing expansive protections to the most vile and hateful expression, a rigorous examination of each approach reveals that such clear-cut characterisations fail to effectively capture the nuances and complexities of each approach. These nuances and complexities warrant attention in debates regarding the application of existing regulatory frameworks to online expression as well as the development of new frameworks targeting this type of speech.

Finally, sometimes lost in debates over the regulation of 'hate speech' are the other tools available to governments to combat rising level of societal hate that do not threaten freedom of expression. These include strengthening anti-discrimination legislation, increasing the capacity of public institutions to tackle intolerance and discrimination, public information and education campaigns directed to combatting negative stereotypes and discrimination against minority groups, and educational initiatives to promote human rights and diversity.[192] These tools are preferable to proscriptions on expression because they target the root causes of societal hate and, as a result, may be more effective than targeting individual speakers. They also engage states in the important work of addressing the systemic and cultural forces that

[192] See, e.g., '"Hate Speech" Explained: A Toolkit (2015 Edition)' (*ARTICLE 19*, 2015) <https://www.article19.org/resources/hate-speech-explained-a-toolkit/> accessed 4 November 2020.

drive inequality and discrimination. These tools are increasingly important in the digital age, as hateful expression proliferates on online platforms and false narratives and conspiracy theories targeting vulnerable and marginalised populations take centre stage in public debates concerning important issues, including the COVID-19 pandemic. A significant challenge here becomes how to contend with situations in which the state itself is the purveyor of extreme speech, especially in the contemporary information ecosystem when government speech reaches audiences far beyond national and regional borders. Chapter 5 tackles this issue through an examination of the contemporary challenges posed by online disinformation from state actors.

Chapter 4

Online Terrorist-Related Expression: Digital Intermediaries and the New Frontier of Online Regulation[1]

'The powerful have erected their current position usually off the backs of violence – not necessarily their own violence, but the violence of their predecessors – and they can celebrate that violence without fear, because they have the power to control the system. But those who have no power in the culture, those who critique the effect of the exercise of power on them, their rival stories of resistance to oppression, of colonial liberation, are condemned as the celebration of terrorism'.[2]

Introduction

In the digital age, digital intermediaries offer both individuals and governments modern-day soapboxes from which to disseminate ideas and information far beyond the proverbial market square. This chapter explores the current landscape of the regulation of online terrorist-related expression and the increasingly aggressive efforts of states and supranational bodies to regulate digital intermediaries. The terrorist attack in New Zealand in early 2019, which was live-streamed on Facebook and rapidly disseminated to other platforms, served as a stark example of the extent to which individuals and groups committing acts of violence for political ends are increasingly using online platforms to spread vitriolic content.[3] The attack contributed to the ongoing debate concerning the extent to which intermediaries have moral and ethical obligations to identify and remove terrorist-related content from their platforms, as well as

[1] Chapter 4 is derived in part from an article written by the author published in the *Journal of Media Law* (copyright: Taylor & Francis): Eliza Bechtold 'Terrorism, the internet, and the threat to freedom of expression: the regulation of digital intermediaries in Europe and the United States' (2020) 12(1) *Journal of Media Law* 13-46.

[2] Conor Gearty, 'The politics of terror' (*Index on Censorship Magazine,* 23 September 2015) <https://www.indexoncensorship.org/2015/09/the-politics-of-terror-conor-gearty/> accessed 20 June 2021).

[3] Kate Lyons, 'Christchurch mosque attacks: suspect charged with "terrorist act"' *The Guardian* (21 May 2019) accessed 20 January 2020.

concerns regarding their ability to do so effectively. While these issues are an important part of a larger conversation regarding the challenges of combatting terrorism in the twenty-first century, they are too often conflated with legal questions regarding the appropriate role of intermediaries in regulating online expression. The latter questions are the focus of this chapter.[4]

Of particular concern is the recent proliferation in Europe of measures aimed at proscribing online speech characterised as 'glorifying' or 'encouraging' terrorism on the basis that such speech may incite future terrorist acts. The justifications for such proscriptions are often couched in terms of the 'unique challenges' posed by the dissemination of this type of expression via online platforms. This reflects a tendency for officials at the national and supranational levels to characterise online terrorist-related expression as a particularly serious and ever-increasing threat to the public. The harm at issue concerns the threat of violence posed by online terrorist-related expression, specifically in relation to future acts of terrorist violence. At the core of these arguments is the notion that there is something particularly harmful about online terrorist-related speech in relation to other types of extreme speech.

This chapter aims to interrogate this assumption by examining questions relating to the appropriate role, if any, of digital intermediaries in regulating online terrorist-related content and the extent to which proponents of human rights should be concerned with the free speech implications of intermediary liability more generally. This examination has implications for the other forms of extreme speech explored in Part II as well as the broader themes of this thesis, as the proper role of intermediaries in policing extreme speech on the internet is an emerging and heavily contested question. The regulation of terrorist-related expression is particularly instructive as it is the focus of new EU legislation that transforms the existing voluntary regulatory model applied to intermediaries into a compulsory one and, thus, may be

[4] As explained in the Introduction, while there is overlap between hate speech and terrorist-related speech in terms of content in that support for terrorist acts or groups often manifests as expressions of racial, religious, and/or ethnic hatred, this thesis treats the latter as a distinct category of expression owing to the fact that the European regulatory framework often distinguishes between terrorist-related expression and other forms of extreme speech, as evidenced by the EU's legislation in this area. See also 'Promotion and protection of the right to freedom of opinion and expression'(*UNGA*, 9 October 2019) UN Doc A/74/486, 13, in which the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression notes that 'States have largely distinguished terrorist and "extremist" content from "hate speech"' and that '[g]overnments that use the term "extremism" in good faith in an online context seem to focus on the problem of the virality of "terrorist and violent extremist ideologies" and seem to have as their goal to counter "extremist" narratives and "prevent the abuse of the internet"'.

used as a model for future regulatory measures. The broader free speech implications of intermediary liability and regulatory sanctions are explored in Chapter 6.

Examining these issues through a comparative lens elucidates the potential implications for freedom of expression that may result from placing legal obligations on intermediaries to police their platforms for extreme speech, including whether generalised and tenuous links between expression and harm are sufficient grounds for proscribing online speech and the extent to which government regulation of online expression via private actors threatens free speech rights. While Europe is adopting increasingly proactive measures to proscribe terrorist-related expression by, inter alia, introducing new frameworks that coerce intermediaries into removing such content from their platforms, the United States is, for the moment, preserving existing frameworks that confer sweeping immunity upon intermediaries for the expression of third-party users in order to encourage widespread and unfettered access to online information. Yet, recent developments suggest that even in the United States, the legislative branch is contemplating stripping existing protections from intermediaries as well as regulatory measures aimed at online expression.

This chapter proceeds in four parts. Part I provides the necessary backdrop for analysing the present questions by examining the role of threat inflation in shaping public perceptions in the United States and Europe regarding the nature and threat of terrorist violence in these regions. Part II analyses the current European approach, which is mainly comprised of a collection of voluntary initiatives, and the shift to a framework that places compulsory obligations on intermediaries to regulate content by way of government imposed standards. Part III analyses the approach of the United States, which immunises intermediaries from liability for the content of third-party users and examines recent challenges to intermediary immunity and the potential for change following events such as Russian interference in the 2016 presidential election. Ultimately, this examination leads to the conclusion that compulsory regulation of intermediaries creates significant dangers to the exercise of free speech, the effects of which ripple far beyond the terrestrial borders of those jurisdictions engaging in such regulation. Thus, efforts to regulate intermediaries should be rigorously interrogated to ensure compliance with relevant free speech principles.

I. The Role of Threat Inflation in Creating Exaggerated Perceptions of the Dangers Posed by Terrorism in the United States and Europe

There is broad consensus in the United States and Europe that terrorism poses a significant and increasing threat to safety and security, and governments in these regions use the threat of terrorist violence to justify increasing proscriptions on human rights, including privacy, due process, and freedom of expression. As discussed in Chapter 1, in both the United States and Europe, limitations on fundamental rights must be proportionate to legitimate government interests.[5] The extent to which government efforts to protect the public from terrorism may infringe on rights is therefore contingent on the nature and severity of the terrorist threat: broadly speaking, the greater the threat, the more latitude the government has in restricting rights. Thus, the only way to determine whether the government overstepped its authority in restricting rights is to assess the extent and gravity of the threat it purports to be addressing.

For decades, scholars have highlighted the extent to which the threat of terrorism is exaggerated by government officials for political purposes, and the resulting impact on the perceptions of the public.[6] This is not to suggest that terrorism does *not* pose a danger to safety and security in the United States and Europe but, rather, that the public's perception of that danger, and of which groups are most likely to commit terrorist acts, is so distorted as to be irrational. This also does not ignore the fact that terrorism is a global phenomenon with significant human rights costs and trans-national implications. However, the focus here is on the irrational perceptions of the threat from terrorism of those living in the United States and Europe, that is, beliefs about the direct threat to their safety and security resulting from possible terrorist attacks in these regions. It is this irrationality, cultivated by government officials, that provides the scope for states to pass increasingly restrictive measures aimed at combatting terrorism that endanger freedom of expression. Accordingly, the extent to which threat inflation shapes and distorts public perceptions of the danger posed by terrorism in the United States and Europe provides the necessary contextual scaffolding for examining government efforts to regulate online terrorist-related content by way of intermediaries.

[5] For example, the ECtHR applies a proportionality test in determining whether an act of a public authority violates Article 10 of the ECHR. In cases involving content or viewpoint-based restrictions on speech, the USSC applies strict scrutiny. See Chapter 1, pp 25 - 28.

[6] See, e.g., John Mueller and Mark G Stewart, '*Terror, Security, and Money: Balancing the Risks, Benefits, and Costs of Homeland Security* (OUP 2011) 175.

In the United States, the evidence shows that fears of terrorist attacks far exceed the actual risk to Americans. For example, while in the period following the 9/11 attacks the probability of an American dying in the United States as a result of terrorism remained steady at one in forty million, forty percent of Americans feared that they or a family member might become a victim of terrorism in 2018.[7] Additionally, while empirical evidence demonstrates that by far the most significant contemporary threat to American national security is from domestic right-wing extremism, a majority of Americans believe that Islamic extremism poses a greater threat.[8] These distorted perceptions are due, in part, to the government's tendency to exacerbate the public's fear of terrorist violence instead of putting the risk of terrorism into perspective.[9] For example, as Mueller and Stewart observe, following 9/11, when the newly formed United States Department of Homeland Security (DHS) opined that '[t]oday's terrorists can strike at any place, any time, and with virtually any weapon,' it should have also pointed out that at the time, anyone living outside of a war zone stood a 1 chance in 85,000 of being killed by terrorism in an eighty year period.[10] Put simply, 'fear of terrorism is as much a function of official communication as it is the result of the attacks themselves'.[11]

Former President Trump's terrorism narratives represented a particularly dangerous iteration of threat inflation that exaggerated the threat of Islamic terrorism for the purpose of building

[7] John Mueller and Mark G Stewart, 'Public Opinion and Counterterrorism Policy (White Paper)' (*The Cato Institute,* 20 February 2018) <https://www.cato.org/publications/white-paper/public-opinion-counterterrorism-policy> accessed 5 May 2019.

[8] Researchers and journalists affiliated with the news site Quartz, using data compiled by the Global Terrorism Database, found that almost two-thirds of terrorist attacks in the United States in 2017 were motivated primarily by racism, Islamophobia, anti-Semitism, homophobia, and/or xenophobia. See Liz Rumero, 'Terrorism Is Surging in the U.S., Fueled by Right-Wing Extremist Ideologies' (*Quartz,* 17 August 2018) <https://qz.com/1355874/terrorism-is-surging-in-the-us-fueled-by-right-wing-extremists/> accessed 25 April 2019. Additionally, every one of the fifty murders in the United States documented by the Anti-Defamation League's Center on Extremism in 2018 was committed by a person or persons with ties to right-wing extremism. See 'A Report from the Center on Extremism: Murder and Extremism in the United States in 2018' (*ADL,* January 2019) <https://www.adl.org/media/12480/download> accessed 13 May 2019. Notwithstanding these facts, in a 2017 survey of American fears, participants reported being most afraid of Islamic extremists/jihadists. See 'Survey of American Fears 2017: Fear of Extremism and the Threat to National Security' (*Chapman University: The Voice of Wilkinson,* 11 October 2017) <https://blogs.chapman.edu/wilkinson/2017/10/11/fear-of-extremism-and-the-threat-to-national-security/> accessed 5 May 2019. Right-wing extremist groups have also proved adept at using social media platforms to disseminate propaganda. In 2016, these groups outperformed ISIS in nearly every social metric on Twitter, including follower counts and tweets per day. JM Berger, 'Nazis v. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks' (*George Washington University Center on Extremism,* September 2016) <https://cchs.gwu.edu/sites/g/files/zaxdzs2371/f/downloads/Nazis%20v.%20ISIS%20Final_0.pdf> accessed 13 May 2019.

[9] Mueller and Stewart 'Terror, Security' (n 7) 176.

[10] ibid.

[11] Geoffrey Edwards and Christopher Meyer, 'Introduction: Charting a Contested Transformation' (2008) 46(1) JCMS 1, 18.

support for controversial policies.[12] For example, an integral part of former President Trump's re-election messaging disinformation[13] regarding the terrorism threat at the United States/Mexico border for the purpose building support for a border wall. In January of 2019, in response to a question about immigration reform, Trump responded '[t]he border is a much more dangerous problem…It's a problem of national security. It's a problem of terrorists. You know, I talk about human traffickers. I talk about drugs. I talk about gangs. But a lot of people don't say - we have terrorists coming through the southern border because they find that's probably the easiest place to come through. They drive right in and they make a left.'[14] These assertions contradicted the findings of the United States intelligence community that there was no credible evidence suggesting that terrorists were actively seeking and/or entering the United States via its border with Mexico during that period.[15] Additionally, former President Trump routinely tweeted following domestic and international terrorist attacks, often using such attacks to justify such controversial policies as the travel ban, which indefinitely suspended the issuance of visas to applicants from certain Muslim-majority countries.[16]

Former President Trump's use of threat *deflation* to play down the increasing threat posed by right-wing extremism in the United States and globally was also an issue of concern. In response to a question from a reporter following the New Zealand attack regarding whether he viewed white nationalism as a rising global threat, President Trump responded 'I don't really.

---

[12] The *New York Times* reports that between January and August of 2019, Trump's re-election campaign posted more than 2,000 ads on Facebook that include the word 'invasion', which form part of a barrage of advertising focused on immigration, a dominant theme of his re-election messaging. See Thomas Kaplan, 'How the Trump Campaign Used Facebook Ads to Amplify His 'Invasion' Claim' *The New York Times* (5 August 2019) <https://www.nytimes.com/2019/08/05/us/politics/trump-campaign-facebook-ads-invasion.html> accessed 12 December 2019.

[13] Extreme speech in the form of disinformation from state actors is the subject of Chapter 5.

[14] 'Remarks by President Trump After Meeting with Congressional Leadership on Border Security' (*Trump White House Archives,* 4 January 2019) <https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-meeting-congressional-leadership-border-security> accessed 25 January 2020.

[15] In a report on terrorism submitted to Congress, the United States Department of State asserted that at the conclusion of 2018 there was 'was no credible evidence indicating international terrorist groups established bases in Mexico, worked directly with Mexican drug cartels, or sent operatives via Mexico into the United States. International supporters and facilitators of terrorist groups such as Hizballah and ISIS are active elsewhere in the Western Hemisphere, and the U.S. southern border remains vulnerable to potential terrorist transit, although terrorist groups are more likely to seek other means of trying to enter the United States.' 'Country Report on Terrorism 2018' (*Department of State*, October 2019) <https://www.state.gov/wp-content/uploads/2019/11/Country-Reports-on-Terrorism-2018-FINAL.pdf> accessed 1 February 2019, 198).

[16] Following a deadly 2017 terrorist attack on a mosque in Egypt, Trump tweeted: '…We have to get TOUGHER AND SMARTER than ever before, and we will. Need the WALL, need the BAN!' 24 Nov 2020. See also Kathy Gilsinan, 'Trump Keeps Invoking Terrorism to Get his Border Wall' (*The Atlantic,* 11 December 2019) <https://www.theatlantic.com/international/archive/2018/12/trump-incorrectly-links-immigration-terrorism/576358/> accessed 10 January 2020. Following a 2017 terrorist attack in London, then-President Trump tweeted 'Another attack in London by a loser terrorist…Must be proactive. The travel ban in the United States should far larger, tougher and more specific – but stupidly, that would not be politically correct!'. 15 Sept 2019.

I think it's a small group of people that have very, very serious problems, I guess.'[17] Regardless of his motivations for minimising the threat posed by right-wing extremism, former President Trump facilitated a discourse that inflated the threat from Islamic terrorist while largely ignoring the much greater threat to American security posed by far-right extremism for the purpose of advancing policies that threatened human rights, both domestically and abroad, under the guise of national security. While the Biden administration is demonstrating a desire to reverse the Trump administration's most egregious counter-terrorism measures, including the travel ban, and is endeavouring to address the threat posed by domestic extremists, the vestiges of four years of xenophobic threat inflation continue to infect American public discourse.[18]

While threat inflation features in the contemporary terrorism discourse in Europe in subtler and less vituperative ways than in the United States, it is utilised for the same ends - to justify increasing restrictions on human rights in the name of protecting the public from harm. Like in the United States, the statistical probability of an EU inhabitant dying from terrorist violence is extremely unlikely.[19] Based on the available evidence, one may reasonably conclude that while there is a tangible threat posed by terrorist violence in Europe, it is not a statistically significant one. Yet, a Special Eurobarometer public opinion survey carried out in the 28 EU countries this same year found that ninety-five percent of respondents regarded security challenges, especially terrorism, as very important.[20] Emphasising the significant disconnect between the actual threat posed by terrorism in Europe and the perceptions of Europeans

[17] Sam Levin, ''It's a small group of people': Trump again denies white nationalism is a rising threat' *The Guardian* (15 March 2019) <https://www.theguardian.com/us-news/2019/mar/15/donald-trump-denies-white-nationalism-threat-new-zealand> accessed 14 April 2019.

[18] One of President Biden's first official acts was to revoke the travel ban (by way of revoking former President Trump's Executive Orders and Proclamations relating thereto). On his first full day in office, President Biden directed his national security team to lead a comprehensive review of the government's efforts to address domestic terrorism, which it recognises as 'the most urgent terrorist threat the United States faces today'. 'FACT SHEET: National Strategy for Countering Domestic Terrorism (*White House Statements and Releases*, 15 June 2021) <https://www.whitehouse.gov/briefing-room/statements-releases/2021/06/15/fact-sheet-national-strategy-for-countering-domestic-terrorism/> accessed 15 July 2021.

[19] In 2017 the statistical probability of an EU inhabitant dying from terrorist violence was 0.0000132 percent. According to Europol's EU 2018 Terrorism Situation and Trend Report, 68 people died in the EU as a result of terrorist violence in 2017. This same year, the EU's Statistical Office estimated that the EU population was 511.8 million. See 'European Union Terrorism and Situation and Trend Report 2018' (*EU Agency for Law Enforcement Cooperation*, 2018) <https://op.europa.eu/en/publication-detail/-/publication/6a2c720a-93a3-11e8-8bc1-01aa75ed71a1/language-en> accessed 1 May 2019, 9; 'First population estimates EU Population up to Almost 512 Million at 1 January 2017. Increase driven by migration' (*EuroStat Press Release*, 10 July 2017) <https://ec.europa.eu/eurostat/documents/2995521/8102195/3-10072017-AP-EN.pdf/a61ce1ca-1efd-41df-86a2-bb495daadbab> accessed 13 May 2019.

[20] 'Europeans' attitudes toward security' (*European Commission*, 2017) Special Eurobarometer 464b <https://europa.eu/eurobarometer/surveys/detail/1569> accessed 30 September 2019.

regarding this threat is not to suggest that government efforts to address terrorism at the national and supranational levels are unwarranted but, rather, that such efforts, and the evidence proffered in support of such efforts, should be rigorously interrogated in circumstances in which human rights are implicated. As discussed below, while European officials are quick to portray terrorism as a serious threat to safety and security in the region that justifies aggressive regulatory reforms, including increased regulation of intermediary platforms, the evidence proffered in support of such claims is often sparse and, in some cases, non-existent.

II. The European Model: Regulation, Regulation, and More Regulation

A. The Definitional Conundrum and the Rise of Glorification Offences

While there are relevant statements concerning terrorism in European and international law, at present, there is no generally accepted definition of the term.[21] While efforts to define terrorism as a legal concept began as early as the 1930s, it was only following the 9/11 attacks that most states began enacting 'terrorism' offences, prompted by an increase in the perceived threat from terrorist violence, new obligations imposed by the UN with respect to the adoption of wide-ranging counter-terrorism measures, gaps in criminal law frameworks, and the conception of terrorism as a unique danger to safety and security.[22] Yet, after nearly 20 years, national laws remain remarkably varied.[23] Additionally, the term terrorism is ideologically and political loaded, and definitional debates reflect doctrinal, ideological, and jurisprudential

[21] See, e.g., UNSC Resolution 1566, which calls upon States to 'to become party…to the relevant international conventions and protocols whether or not they are a party to regional conventions on the matter' and 'to cooperate fully on an expedited basis in resolving all outstanding issues with a view to adopting by consensus the draft comprehensive convention on international terrorism and the draft international convention for the suppression of acts of nuclear terrorism' and UNSC Resolution 1624, which 'calls upon all States to adopt such measures as may be necessary and appropriate and in accordance with their obligations under international law to . . . prohibit by law incitement to commit a terrorist act or acts [and] prevent such conduct'. See also Article 2 of the International Convention for the Suppression of the Financing of terrorism of 1999, which defines an offence thereunder as 'directly or indirectly, unlawfully and wilfully, providing] or collect[ing] funds with the intention that they should be used or in the knowledge that they are to be used, in full or in part, in order to carry…any act intended to cause death or serious bodily injury to a civilian, or to any other person not taking an active part in the hostilities in a situation of armed conflict, when the purpose of such act, by its nature or context, is to intimidate a population, or to compel a government or an international organization to do or to abstain from doing any act'.

[22] See Ben Saul, 'Defining Terrorism: A Conceptual Minefield' (2015) Sydney Law School Legal Studies Research Paper No 15/84 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2664402> accessed 13 May 2019. See also Ben Saul, 'Defining Terrorism: A Conceptual Minefield' in Erica Chenoweth and others (eds), *The Oxford Handbook on Terrorism* (OUP 2019).

[23] ibid. For criticism of the UN's failure to establish an internationally agreed upon definition of terrorism, see Anne F Bayefsky, 'The UN and Incitement' in Laurie R Blank and Anne F Bayefsky (eds), *Incitement to Terrorism* (Brill Nijhoff 2018).

arguments about, inter alia, who is entitled to exercise political violence, for whom, and why.[24] Thus, while there is a general consensus at the international level that terrorism is 'criminal violence intended to intimidate a population or coerce a government or international organisation,' significant disagreements remain, including whether exceptions should be given for 'just' causes such as liberation and rebellion, and whether state violence may constitute terrorism.[25]

This gap in the legal framework at the regional and supranational levels created space for states to define terrorism and related terms, including 'terrorist offence', in increasingly expansive terms.[26] Since 9/11, the UN has issued repeated warnings concerning the human rights implications associated with the absence of a global definition of terrorism, stressing the dangers to freedom of expression resulting from imprecise and overly broad definitions.[27] Similar warnings from international human rights NGOs emphasise the ways in which vague and overly broad definitions ensnare individuals with no connection to terrorism.[28] Of particular concern from a free speech perspective is the recent proliferation throughout Europe of 'glorification' offences, which include offences that criminalise 'glorification' as well as 'advocacy', 'apology' and 'encouragement' of terrorism on the basis that such expression may incite future terrorist acts. While international law proscribes incitement to terrorism, these offences lack the necessary element of intent and/or probability that such expression will lead to violence.[29] Generally speaking, such offences do not require intent to commit a crime, any direct link with an act of terrorism, or even any likelihood that such an act might subsequently occur. As such, they are invariably overly broad and imprecise, and capture expression that falls well outside traditional notions of incitement.

[24] Ben Saul, *Defining Terrorism in International Law* (OUP 2010) 3 - 4.
[25] ibid 4.
[26] Acknowledging the threat to freedom of expression associated with the increasingly expansive counter-terrorism measures adopted by Member States, one of the main priorities of the Council of Europe's Committee on Counter-Terrorism for 2018 to 2019 is to examine the feasibility of reaching agreement on a pan-European legal definition of 'terrorism' for the Council of Europe Convention on the Prevention of Terrorism (Warsaw) CETS No.196.
[27] See, e.g., 'Human Rights Council Report on the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms high countering terrorism' (*UNGA,* 1 March 2019) UN Doc A/HRC/40/52.
[28] See, e.g., 'Statement of Amnesty International', UN High Level Conference of Heads of Counter-Terrorism Agencies of Member States (*Amnesty International*, 29 June 2018) <https://www.un.org/counterterrorism/ctitf/sites/www.un.org.counterterrorism.ctitf/files/S4-Amnesty-International.pdf> accessed 12 February 2019.
[29] ibid.

Concerns regarding the dangers posed by glorification-related offences extend far beyond the protestations of so-called free speech absolutists. In 2018, the Commissioner for Human Rights for the Council of Europe described the proliferation of such offences as a 'disturbing trend', warning of the dangers resulting from the use of overly broad and vague terminology within counter-terrorism laws and emphasising that '[v]iolence and the threat to use violence with the intention to spread fear and provoke terror is the defining component of the concept of 'terrorism''.[30] However, the Council of Europe's Convention on the Prevention of Terrorism defines 'terrorist offence' in much broader terms than the Commissioner recommends, encompassing any of the offences within the scope of and defined in the eleven treaties listed in the Appendix thereto.[31] Moreover, the Convention defines 'public provocation to commit a terrorist offence' as the 'distribution, or otherwise making available, of a message to the public, with the intent to incite the commission of a terrorist offence, where such conduct, whether or not directly advocating terrorist offences, causes a danger that one or more such offences may be committed'.[32]

As with glorification-related offences, the Council of Europe's definition is so broad as to potentially capture speech that neither advocates the commission of terrorist acts nor causes a likelihood that such acts may subsequently occur. Indeed, establishing the mere danger of a particular result is a lower threshold than establishing the likelihood of such a result. However, while the Council of Europe may be taking a 'do as I say, not as I do' approach to the problems

[30] 'Misuse of anti-terror legislation threatens freedom of expression' (*Council of Europe Human Rights Comment*, 12 April 2018) <https://www.coe.int/en/web/commissioner/-/misuse-of-anti-terror-legislation-threatens-freedom-of-expression> accessed 1 April 2019. See also the Joint Declaration on Freedom of Expression and Responses to Conflict Situations (submitted by the UN Special Rapporteur on Freedom of Opinion and expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression and the African Commission on Human and People's Rights Special Rapporteur on Freedom of Expression and Access to Information), which argues that 'states should refrain from applying restrictions relating to "terrorism" in an unduly broad manner. Criminal responsibility for expression relating to terrorism should be limited to those who incite others to terrorism; vague concepts such as "glorifying", "justifying" or "encouraging" terrorism should not be used'. <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=15921> accessed 13 May 2019.

[31] These are: the Convention for the Suppression of Unlawful Seizure of Aircraft, the Convention for the Suppression of Unlawful Acts Against the Safety of Civil Aviation, the Convention on the Prevention and Punishment of Crimes Against Internationally Protected Persons, Including Diplomatic Agents, the International Convention Against the Taking of Hostages, the Convention on the Physical Protection of Nuclear Material, the Protocol for the Suppression of Unlawful Acts of Violence at Airports Serving International Civil Aviation, the Convention for the Suppression of Unlawful Acts Against the Safety of Maritime Navigation, the Protocol for the Suppression of Unlawful Acts Against the Safety of Fixed Platforms Located on the Continental Shelf, the International Convention for the Suppression of Terrorist Bombings, the International Convention for the Suppression of the Financing of Terrorism, and the International Convention for the Suppression of Acts of Nuclear Terrorism.

[32] CETS No 196. (n 26).

associated with overly broad and imprecise definitions of terrorist-related offences, the concerns of the Commissioner, together with the warnings of the UN and human rights NGOs, demonstrate widespread concern at the regional and international levels regarding the erosion of freedom of expression resulting from the absence of definitional constraints.[33]

At the national level, the UK, France, and Spain, have incorporated glorification-related offences into existing counter-terrorism frameworks. The UK added an 'encouragement of terrorism' offence in 2006 and additional related offences in 2019.[34] While the French and Spanish criminal law frameworks included terrorist-related offences prior to 9/11, both countries amended existing offences following the attacks to reflect harsher penalties for online expression, and have recently experienced a spike in prosecutions for such offences.[35] These countries offer numerous illustrations of the absurd and dangerous consequences of government enforcement of such overly-restrictive measures. Amnesty International reports that those prosecuted under Spain's 'glorification of terrorism' offence, added to Article 578 of the Penal Code following the *Charlie Hedbo* attacks in January of 2015 (and for which online expression is considered an aggravating factor) include artists, rappers, puppeteers, and a college student who re-tweeted a historical joke about a prime minister killed by the ETA (Basque Homeland and Liberty) in 1973.[36] Violations of France's 'apology of terrorism' offence, amended in 2014, are punishable by up to five years in prison and a fine of up to 75,000 Euros (increased to seven years in prison and a fine of up to 100,000 Euros if the offence occurs online).[37] Between 2014 and 2016, the number of persons sentenced for this offence rose from 3 to 306, including the conviction of a vegan activist for a Facebook post applauding the death of a butcher in a terrorist attack.[38]

B. A Changing Landscape: The Shift to a Compulsory Framework for Regulating Online Terrorist-Related Content

[33] This mirrors the concerns regarding imprecise or non-existent definitions for 'hate speech' and disinformation discussed in Chapters 4 and 5, respectively, all of which merit demanding scrutiny.
[34] See Terrorism Act 2006 and Counter-Terrorism and Border Security Act 2019, respectively.
[35] Human Rights Comment (n 30).
[36] 'Tweet...If You Dare: How Counter-Terrorism Laws Restrict Freedom of Expression in Spain' (*Amnesty International*, 2018) <https://www.amnesty.org/download/Documents/EUR4179242018ENGLISH.PDF> accessed 15 April 2019. Amnesty International further reports that the vast majority of glorification prosecutions in Spain relate to disbanded or inactive domestic armed groups, namely ETA and GRAPO (First of October Anti-Fascist Resistance Groups), which pose neither an imminent nor significant threat to Spanish national security (the ETA declared a permanent ceasefire in 2017; GRAPO has been inactive since 2007).
[37] Human Rights Comment (n 30).
[38] ibid.

The underlying justifications for glorification-related offences form part of a larger terrorism discourse in Europe that is laden with rhetoric concerning the purported threat to safety and security posed by the dissemination of terrorist-related speech on the internet. By way of example, in 2018, the EU Commissioner for Migration, Home Affairs and Citizenship opined that '[m]any of the recent attacks in the EU have shown how terrorists misuse the internet to spread their messages.'[39] Such broad pronouncements reflect the tendency of European officials at the national and supranational levels to characterise online terrorist-related expression as a particularly egregious and ever-increasing threat to regional safety and security. Conspicuously absent from the justifications proffered by the EU and others for placing increasing responsibilities on intermediaries to police terrorist-related content on their platforms is any empirical evidence demonstrating a causal link between online terrorist-related speech and the commission of subsequent terrorist acts. Indeed, while there is no shortage of assurances that increasing restrictions on freedom of speech are necessary in order to combat the harm resulting from online terrorist-related content, no support for the conclusion that such harm actually exists is proffered. Rather, the resulting harm is simply assumed.[40] This is not to suggest that such support does not exist but, rather, that there should be an onus on the government to present relevant and compelling evidence, rather than general conclusions, in support of sweeping reforms to existing regulatory frameworks that implicate fundamental human rights.[41]

---

[39] 'State of the Union 2018: Commission proposes new rules to get terrorist content off the web' (*European Commission Press Release,* 2 September 2018) <http://europa.eu/rapid/press-release_IP-18-5561_en.htm> accessed 3 April 2019.

[40] For example, in April of 2019, the UK government published the Online Harms White Paper (White Paper), which outlines a new framework with respect to the government's efforts to regulate the online content, with a particular focus on terrorist-related expression. The White Paper references a single source in support of the contention that 'all five terrorist attacks in the United Kingdom during 2017 had an online element, and online terrorist content remains a feature of contemporary radicalisation'. This source is a speech by Minister of Parliament, Amber Rudd, at the 2018 Digital Forum in San Francisco. No information regarding the content of the speech is provided. Rather, the speech is merely referenced in a footnote ('Speech at Digital Forum, San Francisco by the Rt Hon Amber Rudd, 13 February 2018.'). 'Online Harms White Paper' (*HM Government*, April 2019) CP 57, 14 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf> accessed 2 May 2019.

[41] With that said, the available evidence suggesting a causal link between online terrorist-related content and subsequent terrorist acts is weak. For example, in its analysis of the draft version of TERREG, the TGW expressed several concerns. One of the TWG's chief concerns was that it provided insufficient evidence demonstrating a causal link between terrorist acts and 'terrorism content'. In particular, it stressed that while there is evidence that the internet provides a platform for terrorists to effectively disseminate their motivations for committing terrorist acts, the 'available evidence also shows that radicalization tends to occur primarily as a result of offline rather than online dynamics' and that 'radicalization may as well be caused by consumption of daily news (including

1. The EU's Shift to a Compulsory Framework for Regulating Digital Intermediaries

While the EU's influence on the manner in which counter-terrorism policy is conducted at the state level is limited by the principles of sovereignty and subsidiarity outlined in the Introduction[42] it has, in recent years, shown an increased concern regarding the use of intermediary platforms for the dissemination of terrorist-related content.[43] To address the supposed harm resulting from online-terrorist related content, the EU and other European supranational bodies have adopted a framework of ostensibly cooperative initiatives, under which many intermediaries make efforts to detect and remove terrorist-related expression. For example, in 2015, the Commissioner for Migration, Home Affairs and Citizenship launched the EU Internet Forum for the purpose of addressing the 'misuse of the internet by terrorist groups'.[44] It is a voluntary initiative comprised of the EU Home Affairs Ministers, the internet industry and other stakeholders aimed at reducing online terrorist-related content.[45] Underlying this voluntary framework is the principle that intermediaries have societal obligations to protect users by preventing the misuse of their platforms by third parties.[46] Notwithstanding the divergence at the national level regarding what type of terrorist content

---

coverage of terrorist acts).' Accordingly, due to the 'weak evidence for a causal link between terrorism offences and terrorism content', the TWG recommended narrow definitions of material to be targeted by counter-terrorism regulations, namely material that causes an actual risk of and/or imminent harm. Joris van Hoboken, 'The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications (*TWG*, 3 May 2019) <https://www.ivir.nl/publicaties/download/TERREG_FoE-ANALYSIS.pdf> accessed 11 March 2020).

[42] See Javier Argomaniz, Oldrich Bures and Christian Kaunert, 'A Decade of EU Counter-Terrorism and Intelligence: A Critical Assessment' (2015) 30(2-3) Intelligence and National Security 191.

[43] See Introduction, pp 10 - 12. See also Joan Barata, 'New EU Proposal on the Prevention of Terrorist Content Online: An Important Mutation of the E-Commerce Intermediaries' Regime (White Paper)' (2018) (*The Center for Internet and Society at Stanford Law School*, 12 October 2018) <http://cyberlaw.stanford.edu/publications/new-eu-proposal-prevention-terrorist-content-online-important-mutation-e-commerce> accessed 10 March 2019. See also Oldrich Bures, *EU Counterterrorism Policy* (Routledge 2011); Wouter van Ballegooij and Piotr Bakowski, 'The Cost of Non-Europe in the fight against terrorism' (*European Parliamentary Research Service*, May 2018) <https://www.europarl.europa.eu/RegData/etudes/STUD/2018/621817/EPRS_STU(2018)621817_EN.pdf> accessed 5 December 2018.

[44] 'Fighting Terrorism Online: Public-private sector cooperation as important as ever at the fourth EU Internet Forum (*European Commission Statement*, 5 December 2018) <https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_18_6681> accessed 10 February 2019.

[45] See 'EU Internet Forum: Civil Society Empowerment Programme', European Commission <https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network/civil-society-empowerment-programme> accessed 8 February 2020.

[46] See, e.g., Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online (C/2018/1177) [2018] OJ L 63/50 ('[i]n light of their central role and the technological means and capabilities associated with the services that they provide, online service providers have particular societal responsibilities to help tackle illegal content disseminated through the use of their services.' (para 2). This framework includes the Commission Work Programme for committing to continue to promote cooperation with social media companies to detect and remove terrorist and other illegal content online (2018); Commission Communication entitled 'Tackling Illegal Content Online: towards an enhanced responsibility of online platforms' (2017); Council of Europe Convention on the Prevention of Terrorism (Warsaw) (n 32).

the law may proscribe consistent with the right to freedom of expression, there exists a broad consensus among EU Member States in favour of the enactment of legislative and regulatory measures at the supranational and national levels for the purpose of the swift detection and removal of online terrorist-related content.[47]

In April of 2021, the EU enacted a regulation 'on addressing the dissemination of terrorist content online (TERREG) to address the 'misuse of [intermediaries] for terrorist purposes' and to 'contribut[e] to public security across the Union'.[48] TERREG represents the shift to a compulsory model of regulating digital intermediaries in much of Europe, which is predicated on the notion that there is something unique about this category of expression that warrants the creation of compulsory frameworks. While the European Commission targets other forms of harmful online activity, including 'hate speech' and child sexual abuse by way of voluntary regulatory frameworks, this new legislation singles out terrorist-related content for compulsory regulation.[49] The Commission's rationale for this shift, based on generalised references to its Recommendation on measures to effectively tackle illegal content online and the calls of the European Council, is that the online dissemination of terrorist-related content represents a particularly egregious threat to regional safety and security that warrants aggressive measures that are not required to adequately address other forms of harmful online expression.[50] In justifying the shift to a compulsory framework for the regulation of intermediaries, the EU pointed to the purported limitations of voluntary efforts, including that not all intermediaries have chosen to participate, and argues that the overall progress is insufficient to adequately address the existing threat.[51]

TERREG places unprecedented obligations on intermediaries including, inter alia, the removal of terrorist content within one hour of receiving a removal order from a Member State, a duty of care obligation to ensure that platforms are not used for the dissemination of terrorist content and, depending on the circumstances, an obligation to take proactive measures to better protect their platforms.[52] Additionally, obligations are placed on Member States to institute financial

[47] Barata (n 43) 2.
[48] Council and Parliament Regulation (EU) 2021/784 of 29 April 2021 on addressing the dissemination of terrorist content online [2021] OJ L172/79 (TERREG), para 1.
[49] See 'Staff Working Document Impact Assessment Accompanying the document: Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' COM (2018) 640 final, 1-2.
[50] ibid 7.
[51] ibid 25.
[52] TERREG (n 48).

penalties for failure to comply with removal orders.[53] The definition of 'terrorist content', which applies to removal orders, encompasses material that 'directly or indirectly, such as by the glorification of terrorist acts, advocates the commission of terrorist offences, thereby causing a danger that one or more such offences may be committed'.[54] This definition reflects the European trend of adopting increasingly broad and vague definitions of glorification-related offences.

The UN[55] and the EU Agency for Fundamental Rights[56] addressed concerns over some of these provisions in their respective assessments of the draft Regulation. These concerns may be divided into two general categories. The first category relates to the lack of judicial oversight and transparency in the removal order process; the second category concerns the danger of over-regulation given the onerous obligations placed on intermediaries, including enforcement of overly broad and vague definitions of 'terrorist content.' Regarding the first category, TERREG does not require the participation of an independent judicial authority in a Member State's decision to issue a removal order with respect to assessing necessity and proportionality.[57] Additionally, the one hour take-down rule makes it impracticable for an intermediary to effectively challenge a removal order at the relevant time. Regarding the second category the UN, in particular, warns that the extremely short timeline for removal (twenty-four hours) and the threat of significant penalties are likely to incentivise platforms to err on the side of caution and remove content that is legitimate and/or lawful.[58] Additional concerns include that compulsory obligations may enable private actors to remove content that state actors could not restrict consistent with their obligations under domestic and international human rights law, thereby creating the potential for an 'escape route' from human rights oversight.[59]

---

[53] ibid Article 18.
[54] ibid Article 2(7)(a).
[55] See David Kaye, Joseph Cannataci and Fionnuala Ní Aoláin, 'Letter to the European Union' (7 December 2018) OL OTH 71/2018 <https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24234> accessed 9 April 2019.
[56] 'Proposal for a Regulation on preventing the dissemination of terrorist content online and its fundamental rights implications: Opinion of the European Union Agency for Fundamental Rights' (*Publications Office of the European Union*, 2019).
[57] Kaye (n 55) 6. Regarding the issuance of removal orders, TERREG permits Member States to 'decide on the number of competent authorities to be designated and whether they are administrative, law enforcement or judicial'. See TERREG (n 48) para 35.
[58] ibid (Kaye).
[59] ibid 9.

2. The ECtHR's Approach to Adjudicating Challenges to National Glorification-Related Offences

In adjudicating challenges to the application of increasingly broad glorification-related offences enacted at the national level, the ECtHR often holds that such interferences with expression either do not violate Article 10 or that applications are inadmissible per Article 17 on the basis that the expression at issue is *ratione materiae* with the provisions of the Convention.[60] As discussed in Chapter 1, the Court applies Article 17 when speech is directed against the Convention's underlying values, which include tolerance, social peace, and non-discrimination, and in circumstances in which the speaker attempts to rely on the Convention to engage in an activity or perform acts *intended* to destroy the rights and freedoms enshrined therein – regardless of the likelihood of any such effect.[61] In such cases, the Court may find an individual's claim that a government act interfered with his or her Article 10 rights 'inadmissible'. Thus, no Article 10 analysis is required as the rights of the Convention do not attach at all to the contested speech. The Court directs that Article 17 should only be used if is 'immediately clear' that the speech at issue sought to 'deflect [Article 10] from its real purpose by employing to the right to freedom of expression for ends clearly contrary to the values of the convention'.[62] It has interpreted Article 17 to include certain categories of extreme speech, including support for terrorist activity, regardless of whether such speech is directed to inciting terrorist acts.[63] This has allowed space for states to pass expansive restrictions on speech that are increasingly attenuated from traditional notions of harm and incitement.[64]

As highlighted in Chapter 1, the ECtHR's application of Article 17 in Article 10 cases has faced criticism on the basis that it has resulted in categorical exclusions of certain types of speech from Article 10's prima facie protection, which contrasts with the Court's general approach of broad protection for freedom of expression.[65] This practice has also resulted in doctrinal uncertainty as there is no clear test for determining the factual circumstances in which

---

[60] See, e.g., See *Gürbüz and Bayar v Turkey*, App No 8860/13 (ECtHR, 23 July 2019); *ROJ TV A/S v Denmark* App No 24683/14 (ECtHR, 24 May 2018); *Leroy v France*, App No 36109/03 (ECtHR, 10 February 2008) para 43. See also Chapter 1, pp 41 - 45.
[61] See ibid (Chapter 1) pp 41 - 42.
[62] *Ibragimov and others v Russia*, App nos 1413/08 and 28621/11 (ECtHR, 4 February 2019). See also ibid p 42.
[63] *Roj* (n 60).
[64] Human rights penality in the context of 'hate speech' is examined in Chapter 3.
[65] See Chapter 1, p 43 - 45. See also Jonathan Horowitz, 'Case Watch: Europe's Broad View on Acceptable Limits to Free Speech,' (*Open Society Justice Initiative,* 26 April 2013) <https://www.justiceinitiative.org/voices/case-watch-europes-broad-view-acceptable-limits-free-speech> accessed 10 December 2020.

the Court will apply Article 10 versus Article 17.[66] Because a fair proportion of terrorist-related expression may also constitute 'hate speech' against particular religious or racial groups, this tendency makes the application of Article 10 to online terrorist-related speech even more uncertain.[67]

Moreover, while the Court is consistent in its application of the principle that interferences with expression explicitly directed to inciting violence do not violate Article 10, there is less consistency in other types of glorification-related cases.[68] For example, in *Leroy v. France*, the Court held that no violation of Article 10 occurred in the criminal conviction of a man for publishing a drawing representing the 9/11 attack on the World Trade Center with the caption 'we all dreamt it…Hamas did it' notwithstanding the fact that there was no evidence that the applicant intended to incite violence or that his speech increased the risk of subsequent violence.[69] Rather, the Court's decision rested on the rationale that, through his choice of language, the applicant had expressed moral support for the perpetrators of the attacks and, in so doing, expressed approval of the violence and diminished the dignity of the victims.[70] Thus, the Court failed to distinguish between speech directed to inciting violence and speech that merely praises acts of violence.

In 2019, the ECtHR reiterated that Article 10 provides little scope for restrictions on political speech or on debate of questions of public interest 'where the views expressed do not comprise incitement to violence – in other words, unless they advocate resources to violent action or bloody revenge, *justify the commission of terrorists offences in pursuit of their supporter's goals* or can be interpreted as likely to encourage violence by expressing deep-seated and irrational hatreds towards identified persons'.[71] However, the Court does not elaborate on how expression that justifies the commission of terrorist offences constitutes incitement to violence. Indeed, there is arguably a broad range of speech that may be reasonably characterised as

[66] See, e.g., *Pastörs v Germany*, App no 55225/14 (ECtHR, 3 October 2019) para 37 (holding that in cases concerning Holocaust denial, whether the Court applies Article 17 or Article 10 'is a decision taken on a case-by-case basis and will depend on all the circumstances of each individual case'.).
[67] See n 4.
[68] See David J Harris and others, *Harris, O'Boyle, and Warbrick's Law of the European Convention on Human Rights* (4th edn, OUP 2018) 607.
[69] *Leroy* (n 60) 43.
[70] ibid. For additional criticism of *Leroy* and the ECtHR's jurisprudence regarding incitement in glorification-related offences, see Stefan Sottiaux, 'Leroy v France: Apology of Terrorism and the Malaise of the European Court of Human Rights' Free Speech Jurisprudence' (2009) EHRLR 415.
[71] *Alekhina v Russia* (2019) 68 EHRR 1, para 260 (emphasis added). While the facts of this case concerned hooliganism and extremism rather than terrorism, it provides relevant insight into the ECtHR's approach to political speech that takes the form of justifying terrorism.

justification or apology for terrorism that falls well short of traditional conceptions of incitement. For example, the expression for which a vegan activist was convicted under France's 'apology for terrorism' offence was a post on social media concerning the butcher's death in which she stated '[i]t shocks you that an assassin is killed by a terrorist? Not me, I have zero compassion for him. There is justice after all.'[72] While these statements may fairly be characterised as offensive and tactless, there is no reasonable basis upon which to argue that they created a likelihood of future violence against French butchers.

The recent proliferation of glorification related offences at the national and supranational levels in Europe demonstrates the extent to which Europe, unlike the United States, permits increasingly expansive viewpoint-based restrictions on expression. This tolerance may be attributable to the principles underpinning the European approach explored in Chapter 2, including the attachment of rights and responsibilities to the enjoyment of individual rights, the imposition of positive obligations on the government, and the balancing of the rights of an individual speaker with the rights of others, including the targets of virulent speech.[73] These principles contrast starkly with the American conception of democracy, which conceptualises liberty in negative terms and requires the government to establish that proscriptions on speech promoting violence are limited to expression that, as noted above, is both directed to inciting imminent lawless action and is likely to produce such action.[74]

III. The American Model: Open, Free, and Unfettered (For Now)

Unlike Europe, which is enacting increasingly broad viewpoint-based terrorist-related speech offences that fall well outside of traditional notions of incitement and imposing a new framework that places legal obligations on intermediaries to regulate third party content in conformity with such offences, the United States rejects viewpoint-based proscriptions on expression as presumptively unconstitutional, provides sweeping immunity to intermediaries for the content of third-party users, and is preserving existing frameworks predicated on traditional notions of harm and causation. However, Congress's increasing concern regarding

[72] Human Rights Comment (n 30).
[73] See Chapter 2, pp 77 - 80.
[74] See ibid.

the use of social media to promote and disseminate terrorist violence suggests that these long-standing protections may be at risk.

A. Sweeping Immunity for Intermediaries Under the Communications Decency Act

Section 230 of the Communications Decency Act of 1996 (CDA) established a comprehensive statutory immunity to intermediaries for civil liability resulting from the content of third party users.[75] Section 230(c)(1) provides that '[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.'[76] The immunity applies in circumstances in which the interactive computer service provider is not also an 'information content provider,' which the statute defines as someone who is 'responsible, in whole or in part, for the creation or development of' the offending content'.[77] Thus, a defendant is entitled to immunity under the CDA if it is a provider or user of an interactive computer service, the information for which the plaintiff seeks to hold the defendant liable was information provided by another information content provider, and the complaint seeks to hold the defendant liable as the publisher or speaker of that information.[78] With respect to the third element, Congress sought to encourage intermediaries to screen content without fear of liability, thus overriding the traditional treatment of publishers and distributors under statutory and common law.[79] As a result, the CDA immunity extends to the exercise of a publisher's traditional editorial functions, including deciding whether to publish, withdraw, postpone, or alter content created by a third party.[80]

The CDA reflects a set of policy choices by Congress, including not to deter harmful speech through the imposition of liability on intermediaries on the basis that doing so may lead to a

[75] 47 USC s 230.
[76] ibid.
[77] ibid s 230(f)(3). Courts have interpreted the term 'development' as used in section 230(f)(3) definition of 'information content provider' as referring not only to augmenting the content generally, but also to 'materially contributing' to its alleged unlawfulness, that is, for being responsible for what makes the displayed content allegedly unlawful. See, e.g., *Fair Hous Council of San Fernando Valley v Roommates.Com, LLC*, 521 F3d 1157, 1167-1168 (9th Cir 2008); *Jones v Dirty World Entertainment Recordings, LLC.*, 755 F3d 398, 410 (6th Cir 2014). While a website operator like Google can be both an 'interactive computer service' and an 'information content provider', the relevant issue for immunity is whether the interactive computer service provider acts as an information content provider with respect to the information at issue. *Carafino v Metrosplash.com, Inc*, 339 F3d 1119, 1125 (9th Cir 2003) (internal quotations and citations omitted).
[78] *Klayman v Zuckerburg*, 753 F3d 1354, 1357 (DC Cir 2014) (citing CDA (n 75) s 230c(1)).
[79] See *Zeran v Am Online, Inc*, 129 F3d 327, 330 (4th Cir 1997).
[80] ibid.

chilling effect on online expression.[81] In contrast to the E-Commerce Directive, the CDA immunity applies regardless of whether an intermediary is aware of objectionable content and/or whether such content is removed or disabled. Thus, while there is some degree of overlap with respect to the protections afforded to intermediaries under the CDA and the E-Commerce Directive, the former provides significantly broader protection for intermediaries and is predicated upon fundamentally different policy considerations.

B. Contemporary Challenges to the Communications Decency Act: Potential Chinks in the Immunity Armour?

1. Challenging the CDA in the Courts

In adjudicating terrorist-related claims, courts routinely express a preference for broadly construing the CDA. This is primarily attributable to Congress's express desire to permit the continued development of the internet with minimal regulatory interference and to further First Amendment interests for online expression.[82] The disinclination of courts to narrowly construe the CDA immunity has not deterred relatives of victims of terrorist attacks from bringing civil actions against intermediaries, including Facebook, Twitter, and Google. Plaintiffs in these cases primarily rely on various provisions in the Anti-Terrorism Act (ATA), which provide for direct and secondary liability relating to international acts of terrorism.[83] Regarding the interplay between the ATA and the CDA, courts have rejected arguments that the former impliedly abrogated the latter, holding that Section 230 provides a limited defence to a specific subset of defendants (i.e., intermediaries) against the civil liability imposed by the ATA.[84] Thus, so long as the immunity applies, intermediaries may not be held liable for injuries resulting from international acts of terrorism, even if a plaintiff establishes the statutory requirements for liability.

The ATA provides for direct and indirect liability for damages resulting from an international act of terrorism. The direct liability provision permits any national of the United States injured by reason of an act of international terrorism (or their estate or heirs) to sue for damages flowing

[81] s 230 (n 75).
[82] See *Gonzalez v Google*, 335 FSup3d 1156, 1168 (ND Cal 2018).
[83] 18 USC s 2333. The ATA is part of America's robust statutory counter-terrorism framework, ATA Chapter 113B. The most recent proposed amendment to Chapter 113B is HR 4192, entitled 'Confronting the Threat of Domestic Terrorism Act', which inserts s 2332j: 'Acts of terrorism occurring in the territorial jurisdiction of the United States'.
[84] See *Force v Facebook*, 304 FSupp3d 215, 223 (ED NY 2018).

therefrom.[85] Courts have interpreted the phrase 'by reason of an international act of terrorism' to require a plaintiff to establish proximate causation, that is, a direct relationship between the injuries that the plaintiff suffered and the defendant's acts.[86] Indirect liability claims may only be asserted against a person who aids and abets an act of international terrorism, either by knowingly providing substantial assistance or conspiring with the person(s) who committed the act.[87] In order to prevail on a claim of indirect liability, a plaintiff must establish that the party who the defendant aided performed a wrongful act that caused an injury, that the defendant was generally aware of its role as part of an overall illegal activity at the time that such assistance was provided, and that the defendant knowingly and substantially assisted the principal violation.[88]

Plaintiffs in these cases raise similar, and in some cases identical, arguments for imposing liability on intermediaries as those proffered by European officials. These arguments include that intermediary platforms serve as tools for terrorist groups to encourage and incite violence, to spread propaganda for the purpose of radicalising users, to attract new recruits, and to provide mechanisms to raise funds for future attacks.[89] To date, each of these cases has ended in dismissal at the pleading stage for failure to allege the required causal link between the acts of an intermediary and an act of terrorism, the intermediary's assertion of immunity under the CDA, or both.

With respect direct liability claims, courts consistently reject the argument that generalised allegations that a terrorist was radicalised because of online content satisfy the ATA's causation requirement.[90] In adjudicating these types of claims, courts express concerns regarding allegations of tenuous links between intermediaries and the harm resulting from acts

[85] ATA (n 83) s 2333 (a).
[86] See *Clayborn v Twitter*, 2018 WL 6839754, 7 (ND Cal 2018).
[87] ATA (n 83) s 2333 (d).
[88] *Copeland v Twitter*, 2018 WL 6251384, 7 (ND Cal 2018).
[89] See *Cohen v Facebook*, 252 FSupp3d 140 (ED NY 2017); *Crosby v Twitter, Inc.*, 303 FSupp3d 564 (ED NY 2018); *Cain v Twitter*, 2017 WL 1489220 (SD NY 2007).
[90] See, e.g., *Fields v Twitte*r, 881 F3d 739 (9th Cir 2018) (affirming the dismissal of direct liability claims against Twitter on the basis that the plaintiff failed to plead a direct relationship between Twitter's provision of communication equipment in the form of accounts and direct messaging services to ISIS and the injuries sustained by plaintiff resulting from the murder of two American government contractors by Abu Zaid in Jordan); *Taamneh v Twitter*, 2018 WL 5729232 (ND Cal 2018) (holding that conclusory allegations that shooter was radicalised through social media were insufficient to support a plausible claim of contention of probable cause, especially given that there were no allegations that the shooter viewed specific content on social media related to ISIS); *Pennie v Twitter*, 281 FSupp3d 874 (ND Cal 2017) (holding that absent any factual allegations regarding Hamas postings that attacker allegedly viewed and their relationship to the shooting, the assertions that Hamas radicalised the attacker were both too conclusory to be taken as true and too vague to establish proximate cause).

of terrorism, opining that nothing in the ATA suggests that Congress intended to provide a remedy to every person reached by the 'ripples of harm' that flow from intermediaries provision of communications services.[91] Courts are also troubled by the 'untenable litigation risk' that would result from permitting such remedies given the degree to which communication services are interconnected with modern economic and social life.[92] Claims for indirect liability against intermediaries also invariably fail due to plaintiffs' failure to allege any facts beyond that an intermediary was generally aware that a particular terrorist group used its platform.[93]

While plaintiffs suing intermediaries under counter-terrorism statutes advance creative arguments in efforts to circumvent the CDA immunity, courts remain unpersuaded that the mere provision of services, including neutral tools that are used by third parties to post unlawful content, and the use of algorithms that aggregate user and video data to make content recommendations, transform intermediaries into 'content information providers' with respect to the challenged content.[94] Accordingly, Section 230 applies to 'artfully plead' allegations that implicitly require reference to such content to establish liability or implicate an intermediary's role in publishing or excluding third party content.[95]

Thus, at present, the only way to hold intermediaries civilly liable for terrorist-related content on their platforms is to establish both that the intermediary is not subject to the CDA immunity *and* a direct relationship between the injuries resulting from a particular act of terrorism and the acts of an intermediary. The expansive nature of the CDA immunity and the causation requirements of counter-terrorism legislation reflect the extent to which the American framework differs in both underlying policy considerations and practical effect from the increasingly restrictive and onerous measures proliferating in Europe. However, it is worth

[91] See *Clayborn* (n 86) 8.
[92] See *Fields* (n 90) 744.
[93] *Clayborn* (n 86) 9.
[94] See *Gonzalez* (n 82). The plaintiff in *Gonzalez* alleged that Google 'knowingly provided' its YouTube platform and other services to ISIS, and that ISIS 'embraced and used' YouTube 'as a powerful tool for terrorism,' allowing it 'to connect its members and to facilitate [its] ability to communicate, recruit members, plan and carry out attacks, and strike fear in its enemies'. ibid 1170 - 1171. Plaintiff further alleged that Google 'refuse[d] to actively identify ISIS YouTube accounts' or to make 'substantial or sustained efforts to ensure that ISIS would not re-establish the accounts using new identifiers'. ibid 1171. Claims one through four alleged that Google violated the ATA by permitting ISIS and its supporters to publish harmful material on YouTube, and by failing to do enough to remove that content and the users responsible for posting the material. The court held that plaintiff's claims targeted Google's decisions whether to publish, withdraw, exclude, or alter content, which was the type of activity covered by the CDA immunity. ibid.
[95] *Cohen* (n 89) 156.

noting that these protections apply only to intermediaries and not to individual users. While outside the scope of this paper, the USSC recently signalled a notable shift in its approach to individual free speech cases involving providing material support to terrorist organisations in *Holder v. Humanitarian Law Project* (*HLP*).[96] In this case, United States citizens and domestic organisations interested in providing support for the lawful activities of two designated foreign terrorist organisations,[97] including how to petition the UN and other representative bodies for relief and training members how to use international law to peacefully resolve disputes, sought an injunction to prohibit enforcement of a federal statute that criminalised 'knowingly providing material support or resources to a foreign terrorist organization'.[98]

The Court held that the government could prohibit the plaintiffs' proposed activities without violating the First Amendment. It rejected the plaintiffs' argument that the statute restricted 'pure political speech' on the grounds that only 'material support' was targeted and that independent advocacy or membership in international terrorist organisations is not prohibited as the statute was 'drawn to cover only a narrow category of speech to, under the direction of, or in coordination with foreign groups that the speaker knows to be terrorist organisations'.[99] Despite the peaceful nature of the plaintiffs' proposed activities, the Court accepted the government's conclusion that all contributions to foreign terrorist organization, even for ostensibly benign purposes, further terrorist activities.[100] Highlighting the 'sensitive interests in national security and foreign affairs at stake', the Court found that the government met its burden of demonstrating that the statute served the interest of preventing terrorism, even if those providing support intend to promote an organisation's nonviolent ends.[101] The Court emphasised that its holding 'in no way suggest[ed] that a regulation of independent speech would pass constitutional muster, even if the Government were to show that such speech benefits foreign terrorist organizations'.[102]

---

[96] 561 US 1 (2010).
[97] The PKK and the 'Tamil Tigers', which aim to establish independent states for, respectively, Kurds in Turkey and Tamils in Sri Lanka.
[98] ATA (n 83) s 2339B.
[99] *HLP* (n 96) 21..
[100] ibid 31.
[101] ibid.
[102] ibid 34. While the Court thus stressed the narrowness of its holding in *HLP*, human rights defenders have expressed serious concerns regarding the potential scope of the decision. For in depth analysis of the potential

2. The Legislative Branch and Intermediaries: An Increasingly Disharmonious Relationship

While courts in the United States continue to steadfastly apply the CDA in terrorist-related cases, it appears that the Legislative branch is contemplating a shift in the government's approach to intermediaries with respect to the regulation of online content, suggesting that the days of unfettered and unregulated access in the United States may be numbered. Recent events suggest that Congress, the body responsible for enacting the CDA, may be rethinking its approach to the relationship between regulation and intermediaries. For example, in 2018, Facebook, Google, and Twitter were called to testify before Congress to give evidence on their efforts to combat the spread of online extremist content, including terrorist propaganda, in a hearing entitled 'Terrorism and Social Media: Is Big Tech Doing Enough'? Following the New Zealand attack, Facebook and Google testified before Congress regarding their efforts to detect and address white supremacist content on their platforms. The House Judiciary Committee described the purpose of the hearing as to 'foster ideas about what social media companies can do to stem white nationalist propaganda and hate speech online'.[103]

Democrats in Congress are taking an increasing interest in intermediaries in light of the conclusive evidence that Russia interfered in the 2016 election by, inter alia, using intermediaries such as Facebook and Twitter to wage a sophisticated and insidious campaign to destabilise American democracy[104] and undertook similar efforts with respect to the 2020 elections.[105] Intermediaries are also attracting the attention of Republicans in Congress, albeit,

---

free speech implications flowing from this decision with respect to the terrorist-related expression of individuals, see David Cole, 'The First Amendment's Borders: The Place of *Holder v. Humanitarian Law* Project in First Amendment Doctrine' (2016) 6 Harv L & Pol Rev 147; Eliza Bechtold and Gavin Phillipson, 'Glorification of Censorship? Anti-Terror Law, Speech, and Online Regulation' in Adrienne Stone and Frederick Schauer (eds), *The Oxford Handbook on Freedom of Speech* (OUP 2020); James Weinstein and Ashutosh Bhagwat, 'How the United States Supreme Court mishandled the free speech issue in *Holder v. Humanitarian Law Project*' in Ian Cram (ed), *Extremism, Free Speech, and Counter-Terrorism Law and Policy* (Routledge 2019).

[103] US House Committee on the Judiciary, 'April 9: House Judiciary to Hold Hearing on Hate Crimes & White Nationalism' (3 April 2019) <https://judiciary.house.gov/news/press-releases/april-9-house-judiciary-hold-hearing-hate-crimes-white-nationalism> accessed 14 May 2019.

[104] In a 2017 report, the Central Intelligence Agency, Federal Bureau of Investigation, and National Security Agency jointly stated with 'high confidence' that Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the United States presidential election and that Russia's goals were to undermine public faith in the United States democratic process and to denigrate Hillary Clinton in order to reduce her chances of winning the election. See Central Intelligence Agency, Federal Bureau of Investigation, and National Security Agency, 'Assessing Russian Activities and Intentions in Recent US Elections' (6 January 2017) https://www.dni.gov/files/documents/ICA_2017_01.pdf> accessed 10 May 2019.

[105] By way of example, in October of 2019, Facebook publicly identified instances of 'coordinated inauthentic behaviour' on Facebook and Instagram originating in Russia and Iran, and reported removing 93 Facebook accounts, 17 pages and four Instagram accounts that focused primarily on the US, including local and national

for different reasons. For example, Republican Congressman Louie Gohmert, who currently serves as the Vice Chair of the Judiciary Subcommittee on Crime, Terrorism and Homeland Security, introduced a bill in the 116th Congress to remove the CDA immunity from social media companies that use algorithms to filter third party content.[106] According to Mr. Gohmert, the bill is directed to remedying the discrimination against 'conservative voices' on Facebook, Twitter, and Google.[107] Another Republican Congressman, Devin Nunes, recently sued Twitter for purported damages resulting from allegedly defamatory tweets by a third party (named as a co-defendant in the lawsuit). In direct contravention of the CDA immunity, Nunes alleged that Twitter owed him a legal duty of care in the operation of its platform so as not to cause him harm, which it breached in permitting the challenged tweets.[108] The Complaint further alleged that Twitter was actively endeavouring to 'squelch' Nunes's voice as part of a larger pattern of bias against conservatives.[109]

While it is almost certain that Congressman Gohmert's bill will not survive in the Democrat controlled House of Representatives and Twitter was dismissed from Congressman Nunes' lawsuit in June of 2020 on the grounds of the CDA immunity[110], these partisan antics merit attention as potential harbingers of future battles in the United States Congress regarding the regulation of intermediaries. However, any effort to regulate intermediaries would necessarily implicate the First Amendment, which protects the expression of corporations and other associations, as well as individuals.[111] Indeed, any contemplated changes to the existing

political news and public figures. See Nathaniel Gleicher, 'Removing More Coordinated Inauthentic Behavior From Iran and Russia (*Facebook News,* 21 October 2019) < https://about.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-iran-and-russia/> accessed 25 January 2019.

[106] The full title of the bill is the Biased Algorithm Deterrence Act of 2019 (HR 492).

[107] 'Gohmert Introduces Bill That Removes Liability Protections for Social Media Groups that Use Algorithms to Hide, Promote, or Filter User Content' (*US Congressman Louie Gohmert Press Release,* 20 December 2018) <https://gohmert.house.gov/news/documentsingle.aspx?DocumentID=398676> accessed 25 April 2019.

[108] Nunes's complaint may be accessed here: Fox News, 'Nunes Complaint - 3.18.19' <https://www.scribd.com/document/402297422/Nunes-Complaint-3-18-19> accessed 25 April 2019.

[109] ibid 29. A study released in 2021 by the NYU Stern Center for Business and Human Rights revealed that the claim that social media platforms suppress conservative voices 'is itself a form of disinformation: a falsehood with no reliable evidence to 'support it. No trustworthy large-scale studies have determined that conservative content is being removed for ideological reasons or that searches are being manipulated to favour liberal interests. Even anecdotal evidence of supposed bias tends to crumble under close examination'. Paul M. Barrett and J. Grant Sims, 'False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives' (*NYU Stern Center for Business and Human Rights*, February 2021) 1.

[110] In the order granting Twitter's motion to dismiss, the Judge correctly held that Twitter was immune from Nunes's claim per Section 230. See *Nunes v Twitter*, 105 Va Cir 230 (2020).

[111] *Pac Gas and Elec Co v Pub Utils Comm'n of California*, 475 US 1, 8 (1986). See also *Jian Zhang v. Baidu.com Inc*, 10 FSupp3d 433, 437 (SD NY 2014) (holding that online publishers have a First Amendment right to distribute others' speech and exercise editorial control on their platforms because 'the First Amendment's protections apply whether or not a speaker articulates, or even has, a coherent or precise message, and whether or

framework warrant scrutiny as potentially implicating the foundational doctrinal principles underlying the First Amendment examined in Chapter 1, including that government has no authority to limit expression based on its message, its subject matter, its ideas, or its content.[112] Additionally, USSC doctrine extends the rule prohibiting the government from interfering with the editorial judgments of private speakers on issues of public concern beyond the press to corporations and professional publishers.[113]

Given that the American approach is predicated on the principle that the First Amendment generally prohibits content and viewpoint-based restrictions on expression, it is unsurprising that statements of support for terrorist groups and acts, as well as membership in terrorist groups, are protected speech.[114] Additionally, while proscriptions on expression that incites violence are permitted under the First Amendment, such proscriptions are permissible only if the government establishes that the challenged speech is both directed to inciting *imminent* lawless action and is *likely* to produce such action.[115] Proscriptions on the mere advocacy of the use of force or violence are prohibited, as the mere abstract teaching of the moral propriety or necessity for violence are considered legally distinct from preparing a group for violent action and steeling it to such action.[116] Accordingly, glorification-related offences do not pass constitutional muster under the First Amendment for two primary reasons. First, these offences are, by definition, content (terrorism) and viewpoint (glorification or encouragement) based restrictions on expression. Second, they permit restrictions on expression absent any showing of a likelihood of an imminent unlawful act.

Conclusion

not the speaker generated the underlying content in the first place'); *Davison v Facebook*, 370 FSupp3d 621, 629 (ED Va 2019) (holding that Facebook, as a private entity, has 'the right to regulate the content on its platform as it sees fit'); *Le'Tiejira v Facebook*, 272 FSupp3d 981, 991 (SD Tex 2017) (holding that 'Facebook has a [First Amendment] right to decide what to publish and what not to publish on its platform'); *Miami Herald Publ'g Co. v Tornillo* 418 US 241 (1974) (holding a Florida statute requiring newspapers to provide political candidates with free space to reply to attacks on their character as an unconstitutional intrusion into the function of editors in choosing what material goes into a newspaper and in deciding on the size and content of the paper and the treatment of public issues and officials).

[112] See, e.g., *Ashcroft v ACLU*, 535 US 564, 573 (2002).

[113] See *Hurley v Irish-Am Gay, Lesbian & Bisexual Group of Boston*, 515 US 557 (1995).

[114] *See HLP* (n 96) 42.

[115] *Brandenburg* (n 74) 447. The USSC rarely applies the incitement standard and, on the rare occasions when it does, has never found proscriptions on speech to have met the standard. See Erwin Chemerinsky, *Constitutional Law* (4th edn, Aspen Publishers 2013) 1328; *Am Freedom Defense Initiative v Metro Transportation Auth,* 70 FSupp3d 572, 581 (SD NY 2016).

[116] *NAACP v Claiborne Hardware Co,* 458 US 886, 902 (1982).

It is generally accepted that no right, no matter how fundamental, is beyond the reach of government regulation. The relevant question is whether the government, in restricting rights, strikes the proper balance between freedom and regulation. While there is broad consensus that terrorism poses a significant danger to safety and security in the United States and Europe, threat inflation distorts the public's perceptions of the nature and scope of this danger, providing space for governments to exploit exaggerated fears of terrorist violence to justify ever-increasing proscriptions on rights. In the digital age, intermediaries play a central role as gatekeepers of expression in the public sphere. As a result, government efforts to regulate speech on intermediary platforms, whether through formal mechanisms such as legislation, or informal ones such as 'voluntary' partnerships, should be rigorously scrutinised to ensure that freedom of expression is sufficiency protected. Evidence based assessments of the actual threat from terrorist violence in these regions and the impact of increasingly aggressive regulation of intermediaries on freedom of expression should frame the discourse. Instead, the discourse is dominated by threat inflation and generalised assurances that the increasing regulation of intermediaries leads to increased safety and security.

The objective of this chapter is to elucidate the potential free speech implications of placing legal responsibilities on digital intermediaries to police their platforms. A few themes emerge from this exercise. First, the increasing legal entanglement of governments with intermediaries raises serious concerns regarding the erosion of the free speech protections of users. Second, the ways in which governments approach the question of the proper role of intermediaries in regulating online content is heavily influenced by the ways in which notions of harm and causation are conceptualised and incorporated into existing criminal law frameworks. Finally, the extent to which governments permit regulations of this category of speech often reflects the broader value afforded to freedom of expression in relevant human rights frameworks.

While counter-terrorism measures implicate a range of rights, including privacy, religious liberty, and due process, the ubiquity of online speech makes freedom of expression a particularly attractive target for regulation in the digital age. While Europe is shifting to compulsory frameworks in the name of regional security, the United States is, for the moment, continuing to apply existing frameworks that incorporate traditional notions of harm and causation and immunise intermediaries for the expression of third-party users in order to encourage widespread access to information. These differences reflect profound disagreements regarding the appropriate scope of freedom of expression, as well as the role of

the government in regulating online content. Moreover, while America's hands-off approach to regulating online content has, to date, avoided the difficult free speech questions that flow from placing increasing responsibilities on intermediaries to regulate online content, Europe's increasing entanglement with intermediaries illustrates the extent to which government efforts to regulate online content threaten freedom of expression. It also highlights the ways in which the EU has failed to sufficiently address the legal complexities associated with the shift to a compulsory regulatory framework.

At present, as noted above, there is no persuasive evidence supporting the conclusion that proscriptions on speech that merely praises or supports terrorism leads to terrorist violence. Hence, in light of the recent proliferation of glorification-related measures in Europe, the emerging trend of proscribing expression that lies well outside of traditional conceptions of incitement based on highly attenuated or non-existent empirical evidence, and the lack of clarity and coherence in the European approach to regulating terrorist-related expression, the speech protective approach of the United States appears preferable in this area. Indeed, while there is no strong empirical evidence linking terrorist-related expression with subsequent terrorist acts, there *is* abundant evidence that glorification-related offences imperil free speech and permit governments to restrict unpopular expression that poses no danger to safety and security – all under the guise of combatting terrorism.

Finally, often lost in discussions of intermediary liability in the context of terrorist-related expression are larger questions regarding political disagreements concerning what constitutes terrorism and that the way in which governments respond to this question has profound consequences for who may and may not legally promote the use of violence and for what ends in a given jurisdiction. One of the lessons from Europe as it shifts further away from normative frameworks and traditional notions of incitement, and casts intermediaries in the role of enforcers of government proscriptions on expression, is that such conditions create an ideal environment for governments to overregulate and silence the voices of dissidents while sidestepping free speech protections. Free speech defenders in the United States should pay close attention to what is happening in Europe, as the seemingly sacrosanct protection of political speech in the American free speech framework[117] appears to be at risk within a deeply

[117] It is worth noting here the contemporary attacks at the state level in the United States on unpopular political expression characterised as hateful examined in Chapter 3 (see pp 115 - 123).

divided and partisan political climate in which the legislative branch is placing increasing pressure on intermediaries to regulate online content. Chapter 6 considers the lessons from the EU's aggressive approach to the regulation of terrorist-related expression as well as recent efforts at the national level in the United States to regulate intermediaries in other areas in expounding on the free speech implications of increasingly aggressive efforts to regulate digital intermediaries in the context of extreme speech.

Chapter 5

Online Disinformation: State Actors as Purveyors of Extreme Speech

---

'Lies are often much more plausible, more appealing to reason, than reality, since the liar has the great advantage of knowing beforehand what the audience wishes or expects to hear. He has prepared his story for public consumption with a careful eye to making it credible, whereas reality has the disconcerting habit of confronting us with the unexpected, for which we were not prepared'.[1]

Introduction

This chapter examines the emerging threat posed by extreme speech in the form of online domestic disinformation from state actors and the relevant regulatory frameworks in Europe and the United States. It is informed by Hannah Arendt's reflections concerning the publication of the Pentagon papers[2] by the *New York Times* in 1971, which revealed the secret history of the American decision-making process in the Vietnam War and exposed the efforts of successive presidential administrations to prolong the war while lying to the American people regarding the probability of success. The following year, Arendt shared her reflections in an essay entitled 'Lying in Politics'.[3] These reflections provide a useful springboard from which to consider the ways in which the digital age has radically transformed how information in public discourse is disseminated and received, how and where the public consumes such information, and how governments communicate with citizens, for the purpose of examining the contemporary dangers posed by domestic disinformation from state actors.

[1] Hannah Arendt, *Crises of the Republic: Lying in Politics; Civil Disobedience; on Violence; Thoughts on Politics and Revolution* (1st edn, Mariner Books 1972) 6.
[2] Officially titled the 'Report of the Office of the Secretary of Defense Vietnam Task Force', which was commissioned by then-Secretary of Defense Robert McNamara in 1967.
[3] Arendt (n 1) 4.

At the outset, Arendt observes that lying in politics was nothing new in the latter part of the twentieth century, noting that 'the deliberate falsehood and the outright lie used as a legitimate means to achieve political ends, have been with us since the beginning of recorded history'.[4] However, she argues that the crucial point in the context of the Vietnam War was that the policy of lying was employed almost exclusively for domestic consumption, that is, 'for propaganda at home, and especially for the purpose of deceiving Congress'.[5] Arendt called for careful study of the fact that the Pentagon papers revealed little 'significant news' that was not already available to 'the average reader of dailies and weeklies' and the arguments and revelations it raised had been the subject of longstanding debates in various forums, including radio stations and television shows. For Arendt, that the public had access for years to material that the government endeavoured to keep from it demonstrated the integrity of the press even more persuasively than the way that the *New York Times* broke the story:

> What has often been suggested has now been established: so long as the press is free and not corrupt, it has an enormously important function to fulfil and can rightly be called the fourth branch of government. Whether the First Amendment will suffice to protect this most essential political freedom, the right to unmanipulated factual information without which all freedom of opinion becomes a cruel hoax, is another question.[6]

Nearly fifty years later, the internet has fundamentally transformed the information landscape at the national, regional, and international levels, revealing Arendt's question as particularly prescient, in the United States and around the world. For while it is true that lying in politics is nothing new and that state actors have always enjoyed an outsized influence on public discourse, the digital era provides an abundance of new tools and opportunities for these actors to magnify their voices in unprecedented ways. This coincides with a global trend indicating that contemporary threats to democracy typically come from *within* governments rather than from outside actors.[7] In its 2020 Inventory of Organized Social Media Manipulation, the Oxford Internet Institute (Oxford Disinformation Report) found significant evidence that major governments and political parties across the globe are using social media to disseminate disinformation, and that

[4] ibid.
[5] ibid 14.
[6] ibid 45.
[7] See 'Guide Note of the Secretary-General on Democracy', (*UNSG*) <https://www.un.org/en/pdfs/FINAL%20Guidance%20Note%20on%20Democracy.pdf> accessed 3 June 2021.

social media manipulation of public opinion is an increasing threat to democracies around the world.[8] While social media companies face increasing pressure from governments to aggressively root out and remove disinformation from private actors, the challenges of regulating online disinformation become significantly more complex when the purveyor of disinformation is the government itself.

While online disinformation campaigns waged by foreign states and private actors pose an increasing threat to the democratic order, the focus of this chapter is on the challenges of disinformation originating from *within* governments. With Arendt's question regarding whether the First Amendment has sufficed to protect the right to unmanipulated factual information in mind, the primary objective of this chapter is to examine whether online domestic disinformation from state actors has resulted in the failure of European and American free speech frameworks to adequately protect this right. This examination - like those undertaken in the preceding chapters - is situated within a comparative examination of recent developments in Europe and the United States. The harm with which this chapter is concerned is the damage to democratic norms and institutions that results from disinformation campaigns from state actors. It argues that this type of extreme speech poses a greater danger than the other forms of extreme speech examined in this thesis because the resulting harm imperils democracy itself and, in so doing, puts at risk the ideals and safeguards that protect vulnerable and historically marginalised groups from societal hatred and increases the likelihood of terrorist violence.[9]

This chapter proceeds as follows: Part I provides the relevant definitional framework for the discussion, emphasising the importance of precise definitions in this area and the lack of conceptual precision from scholars as well as governments. Part II contextualises the

[8] Hannah Bailey, Samantha Bradshaw and Philip N Howard, 'Industrialised Disinformation: 2020 Global Inventory of Organized Social Media Manipulation' (*Computational Propaganda Project at the Oxford Internet Institute*, 2020) <https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/> accessed 28 April 2020. The Report reveals that disinformation on social media has become a common strategy for political communication in 76 out of the 80 countries studied, including the United States and many European countries. It reflects research carried out between 2019 and 2020 and draws upon a four-step methodology employed by Oxford researchers to identify evidence of globally organised manipulation campaigns, including a systematic content analysis of news articles on cyber troop activity, a secondary literature review of public archives and scientific reports, and generating country specific case studies and expert consultations.

[9] See Asif Efrat and others, 'Report on the Relationship Between Terrorist Threats and Governance Condition in the European Union' (*RECONNECT*, 29 June 2021) <https://reconnect-europe.eu/wp-content/uploads/2021/06/D11.3.pdf> accessed 1 July 2020, 15: 'Processes of democratic backsliding or outright "autocratization" bring countries into the "danger zone" of substantially increased probability of relatively high numbers of terrorist attacks and casualties.'

discussion by examining the shifting landscape of information consumption in Europe and the United States and the resulting disruption to the gatekeeping function served by traditional media. Part III examines the doctrinal principles relevant to the protection afforded to the expression of state actors in public discourse in European and American free speech frameworks, including immunity doctrines that provide expansive protections to such expression. With respect to Europe, much of the focus of this chapter is on the EU response as, in the context of disinformation, its approach is more developed than the Council of Europe and the ECtHR. Part IV interrogates whether these principles adequately protect the public's right to unmanipulated factual information in the digital age through an examination of recent domestic disinformation campaigns waged by state actors in Europe and the United States. Based on the examinations in Parts I through III, this chapter ultimately concludes that domestic disinformation from state actors is the most dangerous form of extreme speech in the digital age.

I. Defining 'Disinformation' and Why Precision Matters in this Context

Like the other forms of extreme speech discussed in this thesis, 'disinformation' is a politically charged term that presents a definitional conundrum. It is often conflated with related types of speech and is misappropriated for political ends. Competing definitions both between and within jurisdictions add to the lack of conceptual clarity. For example, while the European Parliament's Committee on Civil Liberties reported in 2019 that the EU's interinstitutional terminology database IATE (Inter-Active Terminology for Europe) expressly advises that disinformation should not be confused with misinformation, the European Parliament's resolutions use 'misinformation' and 'disinformation' interchangeably.[10] The American government's Cybersecurity and Infrastructure Security Agency defines disinformation broadly as information that is 'deliberately created to mislead, harm, or manipulate a person, social group, organization, or country'.[11]

[10] Judith Bayer and others, 'Disinformation and propaganda - impact on the functioning of the rule of law in the EU and its Member States (Directorate-General for External Policies)' (*Publications Office of the European Union*, 2019) 25.
[11] 'COVID-19 Disinformation Toolkit' (*CISA.gov,* 2020) <https://www.cisa.gov/sites/default/files/publications/SLTTCOVIDToolkit_FINAL_508.pdf> 30 April 2021.

As emphasised throughout this thesis, precise definitions in the area of speech regulation are important because they often delineate the line between expression that may be lawfully proscribed and that which is immune from government intrusion. Precision is particularly important here because this chapter posits that disinformation from state actors represents a unique and increasingly dangerous form of extreme speech in the digital age. This thesis does not attempt to provide a universal definition of 'disinformation' that is applicable across jurisdictions and in all contexts. Rather, the objective is to adopt and apply a definition that precisely reflects the scope of this discussion. To that end, this chapter adopts the following definition of disinformation: 'verifiably false information created, presented and disseminated to intentionally deceive the public'.[12]

This definition precisely captures the objective of this chapter in two important ways. First, it highlights that the information at issue is verifiably false, that is, its falsity is not something upon which there may be reasonable disagreement. While definitions of disinformation often include 'misleading', this term is excluded here because of the important distinction between information that is misleading and information that is false, the crucial difference being that only the latter - at least in theory - can be objective established.[13] This raises the related point of the challenges of discerning fact from opinion in a 'post-truth' environment.[14] Efforts by state and political actors to obfuscate the important distinction between fact and opinion in public discourse highlight the importance of precise definitions in this context.

Second, unlike misinformation, which is commonly defined as false information that is spread *without* intent to mislead and is often shared because the user mistakenly believes it to be true, disinformation is shared with the *express* purpose of misleading the public.[15] Disinformation is also preferable to 'fake news', which has been appropriated by state and political actors to

[12] This closely aligns with the definition proffered by the European Commission. See 'Tackling online disinformation' (*European Commission*) <https://digital-strategy.ec.europa.eu/en/policies/online-disinformation> accessed 1 July 2021.
[13] See Björnstjern Baade, 'Don't Call A Spade A Shovel: Crucial Subtleties in the Definition of Fake News and Disinformation' (*Verfassungsblog,* 14 April 2020) <https://verfassungsblog.de/dont-call-a-spade-a-shovel/> accessed 20 May 2021.
[14] "Post-truth" was Oxford Dictionaries' Word of the Year in 2016, defined as 'relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief'. This word was chosen because of a spike in frequency in 2016 in the context of Brexit and the American presidential election, and its association with 'post-truth politics.' (*OUP*, 'Word of the Year 2016' <https://languages.oup.com/word-of-the-year/2016/> accessed 20 May 2021).
[15] See Commission (n 12).

dismiss critical coverage and to attack and undermine the press.[16] Another problem with the term 'fake news' is that it frames the problem of disinformation as isolated incidents rather than strategic efforts to deceive the intended audience.[17]

II. The Shifting Landscape of Information Consumption in Europe and the United States

In addition to adopting a precise definition of disinformation, it is important to contextualise this discussion by sketching the current landscape of information consumption in Europe and the United States. This provides the necessary backdrop for the examinations undertaken in Part III of this chapter.

A. The Rise of Social Media as a Source for Political News

Europeans and Americans are increasingly turning to the internet for political news, creating an ideal environment for state actors to weaponise social media and other online platforms to disseminate disinformation to the masses. In a Pew Research Center poll conducted between October 2019 and June 2020, social media was among the most common source of political news for Americans and more popular than network and local television and radio.[18] During this same period, less than thirty-eight percent of any segment of the United States population often relied on print newspapers.[19] Facebook is the most common social media site used for news by Americans, with approximately four-in-ten getting news on this platform[20]

While social media is now a widely used source of news for many Americans, recent data suggests that it is less popular for Europeans. For example, a 2018 survey of seven Western

[16] 'A multi-dimensional approach to disinformation: Report of the independent High-level Group on fake news and online disinformation' (*Publications Office of the European Union*, 2018) <http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271> accessed 1 July 2021. See also Claire Wardle, 'Fake News. It's Complicated' (*First Draft,* 16 February 2018) <https://firstdraftnews.org/articles/fake-news-complicated/> accessed 21 May 2021.
[17] See Claire Wardle and Hossein Derakshan, 'INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making' (Council of Europe Report DGI(2017)19' (*Council of Europe*, 27 September 2017)) <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c> accessed 30 June 2021.
[18] Amy Mitchell and others, 'Americans Who Mainly Get Their News on Social Media are Less Engaged, Less Knowledgeable' (*Pew Research Center,* 30 July 2020) <https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/> accessed 30 June 2021.
[19] ibid.
[20] Abigail W Geiger, 'Key findings about the online news landscape in America' (*Pew Research Center,* 11 September 2019) <https://www.pewresearch.org/fact-tank/2019/09/11/key-findings-about-the-online-news-landscape-in-america/> accessed 16 June 2021.

European countries revealed a public news organisation as the public's primary source for news.[21] While many younger Europeans get some of their news from social media, this mode is more popular in the United States, especially for older adults. In the UK, Sweden and the Netherlands especially, adults are more attached to their public broadcasters.[22] This also contrasts starkly with the United States, where the largest public news outlets, NPR and PBS, rank far lower than many of the country's private news outlets.[23]

As more and more people turn to social media for political news, these platforms become an increasingly popular forum for state actors to communicate with citizens. For example, prior to his suspension from Twitter in January of 2021, former President Trump had over 81 million followers, the most of any world leader at that time.[24] In 2020, ninety-nine percent of United States Senators and ninety-eight percent of House members posted on official social media accounts, and all of these lawmakers posted on Facebook.[25] Collectively, members of the United States Congress sent over five hundred thousand tweets in 2020 alone and were responsible for over two hundred thousand Facebook posts.[26] Between January 2019 and June 2020, thirty-four EU heads of state or government posted over 31,000 tweets.[27]

B. The Precipitous Decline of Public Trust in Traditional Media Outlets and Increasing Political Polarisation: A Uniquely American Problem?

[21] Katerina Eva Matsa, 'Across Western Europe, public news media are widely used and trusted sources of news' (*Pew Research Center,* 8 June 2018) <https://www.pewresearch.org/fact-tank/2018/06/08/western-europe-public-news-media-widely-used-and-trusted/> accessed 15 March 2021. According to Ofcom's 2019 news consumption report, BBC One remains the most popular news source for people living in the UK. The report also revealed that while television remains the most popular way that people access news in the UK, use of social media for news rose from forty-four percent in 2018 to forty-nine percent in 2019 (Jigsaw Research, 'News Consumption in the UK: 2019' (*Ofcom,* 24 July 2019) <https://www.ofcom.org.uk/__data/assets/pdf_file/0027/157914/uk-news-consumption-2019-report.pdf> accessed 2 June 2020.

[22] ibid (Matsa).

[23] Mitchell (n 18).

[24] Hristina Tankovska, 'The world leaders with the most Twitter followers 2020' (*Statista*, 27 January 2021) <https://www.statista.com/statistics/281375/heads-of-state-with-the-most-twitter-followers/> accessed 3 February 2021.

[25] ibid.

[26] ibid.

[27] Ralf Drachenberg and Emily Phillips, 'The Twitter activity of members of the European Council: A content analysis of EU leaders' use of Twitter in 2019-20' (*Publications Office of the European Union*, 8 January 2021) <https://www.europarl.europa.eu/RegData/etudes/STUD/2021/654200/EPRS_STU(2021)654200_EN.pdf> accessed 20 May 2021. See also Giles Brachotte and others (eds), *Tweets from the Campaign Trail: Researching Candidates' Use of Twitter During the European Parliamentary Elections* (Peter Lang AG 2016).

In 2019, researchers from the Reuters Institute for the Study of Journalism at the University of Oxford, at the request of the European Parliament's Panel for the Future of Science and Technology, conducted a literature review examining the effect of news use on polarisation across Europe.[28] Among the key findings that emerged from this review are that there is limited evidence to suggest that increased exposure to news featuring like-minded or opposing views leads to widespread polarisation of attitudes in Europe, and most studies of social media use in Europe have not found evidence of echo chambers.[29] A 2019 report from the European Broadcasting Union showed, for the first time, that more people in Europe tend to trust, rather than distrust, the written press.[30] It also revealed that public service media (PSM) are among the top five most trusted news brands in more than eighty percent of North and Central European markets and in more than sixty percent of those markets, PSM is the number one brand for trusted news.[31]

Comparative research suggests that the United States has much higher levels of partisan news production, consumption, and polarisation than Europe and lower levels of trust in traditional media.[32] Even when accounting for the differences within European states, and the rise in the use of social media as a news source across Europe, the United States' information ecosystem is significantly more polluted, fragmented, and polarised.[33] Additionally, partisan polarisation

---

[28] Richard Fletcher and Joy Jenkins, 'Polarisation and the news media in Europe' (*Publications Office of the European Union*, 2019) <https://op.europa.eu/en/publication-detail/-/publication/914380a0-8e62-11e9-9369-01aa75ed71a1/language-en> accessed 18 May 2021.

[29] Echo chambers are places where the reinforcing effect of media and beliefs drive people to wall themselves off from a wider range of media. See Jay Hmielowski, Myiah Hutchens and Michael Beam, 'Conservatives are more likely than liberals to exist in a media echo chamber' (*London School of Economics Blog,* 22 December 2020) <https://blogs.lse.ac.uk/usappblog/2020/12/22/conservatives-are-more-likely-than-liberals-to-exist-in-a-media-echo-chamber/> accessed 3 February 2021.

[30] 'Trust in Media 2020' (*EBU Research,* June 2020) <https://www.ebu.ch/publications/research/login_only/report/trust-in-media> accessed 25 June 2021.
This report is primarily based on data from the 92nd Standard Eurobarometer, from where the Net Trust Index is deducted. The 92nd wave of the survey was conducted in November 2019 in the 28 EU Member States and the five candidate countries. In the section about trust in news, the Reuters Institute Digital News Report 2019 is used as an additional source, whereas the section on trust and COVID-19 is based on survey data from Global Web Index, Reuters Institute, Ofcom and EBU member organizations.

[31] ibid. Of course, there are notable exceptions to this trend. By way of example, in Hungary - which Freedom House demoted to the status of a 'transitional or hybrid regime' in 2020, noting a methodical dismantling of the institutions and norms sustaining democracy orchestrated by right-wing nationalist governments - trust in traditional media is extremely low, suggesting a highly polarised media environment where consumers are drawn to brands that reflect their political views. Eva Bognar, 'Digital News Report: Hungary' (*Reuters Institute,* 2020) <https://www.digitalnewsreport.org/survey/2020/hungary-2020/> accessed 7 June 2021. See also, Zselyke Csaky, 'Nations in Transit 2021: The Antidemocratic Turn' (*Freedom House,* 2021) <https://freedomhouse.org/report/nations-transit/2021/antidemocratic-turn> accessed 12 May 2021.

[32] ibid. See also Christopher A Bail and others, 'Exposure to opposing views on social media can increase political polarization' (2018) 115(37) PNAS 9216.

[33] Trust in Media (n 30).

in the use and trust of media sources has widened in the United States over the past five years.[34] While Republicans have grown increasingly alienated from most traditional news sources, Democrats' confidence in them remains stable, and in some cases, has strengthened.[35] A 2019 Pew Research Centre survey found that Republicans and Democrats trust 'two nearly inverse news media environments'.[36] Recent research also reveals that while both liberal and conservative American media contribute to polarisation, conservatives tend to trust fewer media sources and are more likely to be part of media echo chambers than liberals because the latter consume a wider range of views.[37] Notwithstanding these marked partisan differences, recent studies suggest overall declining levels of public trust in traditional journalism in the United States. For example, the percentage of Americans reporting they had a great deal or fair amount of trust in traditional journalism dropped from fifty-three percent in 1997 to thirty-two percent in 2016, and a majority of Americans believe major news organizations 'routinely produce false information'.[38]

Recent declines in trust in traditional media by Republicans coincide with a recent increase of illiberalism of the Republican Party over the last generation. In January of 2020, the VDem Institute at the University of Gothenburg released the results of a data set - the 'illiberalism index' - which gauges the extent of commitment to democratic norms a party exhibits before an election.[39] Indicators comprising the index are low commitment to political pluralism, demonization of political opponents, disrespect for fundamental minority rights, and

[34] Pew surveyed 12,043 American adults in October and November of 2019 and asked whether they had heard of or used any of 30 media sources, chosen so that respondents were asked about a range of news media across different platforms. Everyone who took part is a member of the Pew Research Center's American Trends Panel, an online survey panel that is recruited through national, random sampling of residential addresses. See Mark Jurkowitz and others, 'U.S. Media Polarization and the 2020 Election: A Nation Divided' (*Pew Research Center,* 24 January 2020) <https://www.journalism.org/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/> accessed 8 June 2020.
[35] ibid.
[36] ibid.
[37] Hmielowski (n 29).
[38] Daniel M West, 'How to combat fake news and disinformation' (*Brooking Institution*, 18 December 2017) <https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/> accessed 13 June 2021.
[39] The largest ever study of its kind, the illiberalism index highlights shifts and trends within and between political parties since 1970. 665 experts assessed the identity of the political parties in their country of expertise with a vote share of more than five percent in a legislative election between 1970 and 2019 across 169 countries. This generated a dataset of 1,955 political parties across 1,560 elections, which resulted in 6,330 party-election year units within 183,570 expert-coded data points. Coder responses were aggregated using V-Dem's custom-built statistical model to ensure comparability across countries and time. The institute describes the index as 'the first comparative measure of the "litmus test" for the loyalty to democracy.' See Garry Hindle and others, 'New Global Data on Political Parties: V-Party (Briefing Paper No 9)' (*V-Dem Institute,* 26 October 2020) <https://www.v-dem.net/media/filer_public/b6/55/b6553f85-5c5d-45ec-be63-a48a2abe3f62/briefing_paper_9> accessed 3 January 2021, 1.

encouragement of political violence. The illiberalism index reveals that the Republican Party has retreated from upholding democratic norms and that its contemporary rhetoric is closer to authoritarian parties - including AKP in Turkey and Fidesz in Hungary – than to parties in other democracies.[40] Additionally, based on the above-referenced indicators, the Republican Party in 2018 was *far* more illiberal than almost all other governing parties in democracies, with only fifteen percent of governing parties in democracies in the twenty-first century considered more illiberal than the contemporary Republican Party.[41]

Declining levels of trust in traditional media are concerning because it serves the important free speech function of accurately informing individuals regarding matters of public concern and holding those in power to account. While a healthy press system can generally absorb occasional official attacks on the press and sporadic partisan campaigns, public information systems that develop large media networks that routinely spread disinformation and engage in sustained attacks on traditional media organisations create risks to democratic stability.[42] And while the press is not invariably the first democratic institution to be attacked when a country's leadership takes an anti-democratic posture, repression of free media is a strong indicator that other human rights are in danger.[43]

As more people turn to social media and other online sources for political news, the more space becomes available for state actors to wage disinformation campaigns for the purpose of improperly manipulating public opinion. Existing free speech frameworks largely reflect the pre-digital era, in which fewer information sources operated, enabling traditional news organisations to exercise more effective gatekeeping against disinformation.[44] Part III explores these frameworks.

III. The Protection Afforded to Expression from State Actors in European and American Free Speech Frameworks

[40] ibid.
[41] In contrast, the Democratic Party rated only slightly less illiberal than the typical party in democracies and has retained a commitment to longstanding democratic standards. ibid.
[42] W Lance Bennett and Steven Livingston, 'The disinformation order: Disruptive communication and the decline of democratic institutions' (2018) 33(2) EJ Comm 122, 125.
[43] Sarah Repucci, 'Media Freedom: A Downward Spiral' (*Freedom House*, 2019) <https://freedomhouse.org/report/freedom-and-media/2019/media-freedom-downward-spiral> accessed 8 June 2021.
[44] Bennett (n 42)

During the second impeachment trial of former President Trump following the Capitol insurrection on 6 January 2021, his lawyers argued that at the 'Save America' rally, Mr. Trump was simply exercising his First Amendment right as a citizen to engage in political speech.[45] While acknowledging former President Trump's position as an elected official, they argued this fact was irrelevant to the application of the First Amendment to his speech. These arguments highlight the two important doctrinal questions that inform this chapter. First, how do European and American free speech frameworks treat expression from state actors in public discourse? Second, do courts distinguish between state actors speaking in their private and official capacities and, if so, how do such distinctions impact upon the protection afforded to such expression? While protection of statements from state actors, in particular legislators, varies by jurisdiction, immunity for statements made in the legislature and in similar bodies are common in liberal democracies.[46]

While different approaches are used to achieve immunity, its general purpose is always to enable legislative bodies to carry out tasks without undue external interference. Indeed, '[s]cholarly literature, case law, and the practice of most national parliaments and many other representative bodies' generally recognise that 'the protection which immunity affords is indispensable to the operation of democracy'.[47] The European and American approaches with respect to such immunity are discussed below.

A. Regulating the Speech of State Actors in Europe: Immunities Within a Broader Framework that Attaches Duties and Responsibilities to the Exercise of the Right to Freedom of Expression

1. The Scope of Parliamentary Immunity in the EU

At the EU level, members of the European Parliament (MEPs) are granted a special regime of immunities, including absolute immunity for opinions expressed in the performance of their duties, and protection from prosecutions and restrictions of their personal freedom during the

[45] Bruce L Castor, Jr, David Schoen and Michael T van der Veen, 'Trial Memorandum of Donald J. Trump, 45th President of the United States of America' (Integral Text Publishing 2021) 44 - 45.
[46] See Sascha Hardt, 'Parliamentary Immunity in a European context (Directorate General for Internal Policies (Policy Department C: Citizens' Rights and Constitutional Affairs))' (*Publications Office of the European Union,* 2015) <https://publications.parliament.uk/pa/jt199899/jtselect/jtpriv/43/43ap21.htm> accessed 15 May 2021.
[47] ibid 7 (internal quotations omitted).

sessions of the Parliament.[48] MEPs, like members of national parliaments in the majority of EU Member States, enjoy absolute immunity for the opinions expressed and votes cast in the exercise of their duties, and relative immunity from arrest and detention during the sessions of the European Parliament.[49]

Absolute immunity is generally intended to apply to opinions expressed in the premises of the European Parliament.[50] However, it is possible that an opinion expressed outside the European Parliament may amount to an exercise of an MEP's duties so long as it is an assertion amounting to 'a subjective appraisal having a direct, obvious link with the performance of those duties'.[51] The final decision as to whether such an opinion is expressed in the exercise of the MEP's duties pertains to the exclusive jurisdiction of the national courts.[52] While absolute immunity derives exclusively from EU law, and is therefore uniform for all EU Member States, the scope of personal immunity partly depends on the rules applicable to national Members of Parliament.[53] For example, under the Italian constitution, senators enjoy a broad parliamentary immunity, including for statements made outside of parliament.[54]

2. The Scope of Parliamentary Immunity in the Council of Europe and the ECtHR

In June of 2021, the Parliamentary Assembly of the Council of Europe Committee on Legal Affairs and Human Rights[55] (Committee on Human Rights) released a report in response to the question of whether 'politicians should be prosecuted for statements made in the exercise of their mandate'?[56] The authors of the motion underlying the report were 'concerned about the

[48] See Treaty establishing a Constitution for Europe [2004] OJ C310/261, Protocol 7 on privileges and immunities of the European Union.
[49] Roberta Panizza and Eeva Pavy, 'Handbook on the incompatibilities and immunity of the Members of the European Parliament' (*Publications Office of the European Union*, March 2020) 33. The Handbook provides a comprehensive overview of the EU framework for parliamentary immunity.
[50] ibid 12.
[51] See Case C-163/10 *Criminal proceedings against Aldo Patriciello* [2011] ECR I-07565, para 32.
[52] Rosa Raffaelli and Sarah Salome Sy, 'The Immunity of Members of the European Parliament (Directorate-General for Internal Policies, Policy Department C: Citizens' Rights and Constitutional Affairs)' (*Publications Office of the European Union*, 2014) 5. The EU Policy Departments provide in-house and external expertise to support European Parliament committees and other parliamentary bodies in shaping legislation.
[53] ibid.
[54] Marília Crespo Allen, 'Parliamentary immunity in the members states of the European Union and in the European Parliament' (*Legal Affairs (Series W-4),* February 1993).
[55] The Committee promotes the rule of law and defends human rights. It is also responsible for various activities that make it the de facto legal advisor to the Parliamentary Assembly, which is the deliberative body of the Council of Europe, to which parliamentarians are appointed by the national parliaments of the Assembly's Member States.
[56] Boriss Cilevičs, 'Should politicians be prosecuted for statements made in the exercise of their mandate? (Provisional Version)' (*Council of Europe Committee on Legal Affairs and Human Rights*, 7 June 2021).

growing number of national, regional and local politicians prosecuted for statements made in the exercise of their mandate, in particular in Spain and Turkey'.[57] The report highlights the Venice Commission's view that 'the primary purpose of parliamentary immunity lies in the fundamental protection of the parliamentary institution and in the equally fundamental guarantee of the independence of elected representatives, which is necessary for them to exercise their democratic functions effectively without fear of interference from the executive or judiciary'.[58] The report further notes the Venice Commission's view that 'parliamentarians' freedom of speech must be a wide one and should be protected also when they speak outside Parliament'.[59] Such a wide interpretation of freedom of speech should apply, in particular, 'to parliamentarians who belong to the opposition and whose ideas differ strongly from those of the majority'.[60] At the same time, the report highlights that the motion 'insists that hate speech and calls for violence cannot be tolerated, also from politicians'.[61] The report further stressed 'the crucial importance, in a living democracy, of politicians being able to freely exercise their mandates', which 'requires a particularly high level of protection of politicians' freedom of speech and freedom of assembly, both in parliament and when speaking to their constituents in public meetings or through the media'.[62]

As discussed in Chapter 2, while the ECtHR interprets Article 10 as permitting individuals participating in public discourse a degree of exaggeration and provocation, it also imposes a duty not to exercise the right to freedom of expression in an irresponsible or excessive manner.[63] Of primary concern is the maintenance and promotion of the ideals and values of a democratic society, including pluralism and tolerance, for which it is considered appropriate to subordinate individual interests in order to ensure the fair and proper treatment of marginalised and vulnerable groups and to protect such groups from harm.[64] Thus, as duties and responsibilities attach to all persons who exercise the right to freedom of expression, the relevant questions here are whether state actors are subject to *enhanced* duties and responsibilities due to their positions and the resulting influence they wield in public discourse

[57] ibid 5.
[58] ibid.
[59] ibid.
[60] ibid 7
[61] ibid.
[62] ibid 1.
[63] See Chapter 2, p 80.
[64] ibid.

and/or whether there are circumstances in which state actors receive *more* protection for their speech based on their status.

While the ECtHR has opined that persons of influence, including politicians and party leaders, owe a particular responsibility due to their enhanced influence on their followers, it has not proffered a definition or test to assist in identifying who may count as an influential figure and on what basis. In *Willem v. France,* the Court highlighted that a public servant such as a mayor has particular duties and responsibilities, including a 'degree of neutrality and reserve with regard to the territorial community which he represents as a whole'.[65] The status of the applicant as a state actor was also addressed in *Féret v. Belgium*, which found no violation of Article 10 with respect to the conviction of the applicant - chairman of the 'Front National' party and member of the Belgian House of Representatives - for publicly inciting discrimination or hatred following complaints concerning leaflets distributed by the party during election campaigns.[66] The Court opined that it was the 'duty of politicians to refrain from using or advocating racial discrimination and recourse to words or attitudes which are vexatious or humiliating because such behaviour risks fostering reactions among the public which are incompatible with a peaceful social climate and could erode confidence in democratic institutions'.[67]

In other cases, the Court has emphasised the particular importance of freedom of expression for Members of Parliament as 'political speech *par excellence*'.[68] It has also emphasised that while 'freedom of expression is important for everybody' it is especially so for an elected representative who 'represents the electorate, draws attention to their preoccupations and defends their interests'.[69] As such, interferences with the expression of elected representatives 'call for the closest scrutiny'.[70] Notwithstanding these pronouncements, the ECtHR recognises limits to parliamentary privilege. For example, while there is a broad range of circumstances in which the ECtHR will tolerate defamatory statements,[71] it is less permissive of falsehoods that target vulnerable and marginalised communities in the context of parliamentary privilege.

[65] *Willem v France*, App No 10883/05 (ECtHR, 16 July 2009).
[66] *Féret v Belgium*, App no 15615/07 (ECtHR, 16 July 2009).
[67] ibid para 77. See also Tarlach McGonagle, 'The Council of Europe against online hate speech: Conundrums and challenges (Expert Paper)' <https://rm.coe.int/16800c170f> accessed 17 June 2021 (note that the excerpt referenced was translated from French to English by the author as the judgment is only available in French).
[68] See *Pastörs v Germany*, App no 55225/14 (ECtHR, 3 January 2020), para 47.
[69] *Castells v Spain*, App no 11798/85 (23 April 1992) para 42.
[70] ibid.
[71] This includes defamation of politicians. See, e.g., *Lingens v Austria*, App no 9815/82 (ECtHR, 8 July 1986).

In so doing, the Court recognises that certain types of defamation are at odds with the democratic values of the Convention and, as a result, are deserving of 'little, if any protection'.[72] This position derives from the principle that the exercise of freedom of expression, 'even in Parliament', carries with it the duties and responsibilities outlined in paragraph 2 of Article 10.[73]

This principle is illustrated in the ECtHR's decision in *Pastörs v. Germany*, which concerned a sitting Member of Parliament's statements during a Parliamentary session that contained a qualified Holocaust denial. The Court held the applicant's conviction for violating provisions of the German Criminal Code concerning the memory of the dead and of defamation did not violate Article 10 because 'the impugned statements affected the dignity of the Jews to the point that they justified a criminal law response'.[74] While acknowledging that Member States have 'very limited latitude' in regulating the content of Parliamentary speech,' it reasoned that some regulation is necessary in order to prevent forms of expression such as direct or indirect calls for violence.[75] Thus, while the ECtHR recognises a general immunity for statements made by lawmakers, such immunity is not absolute and is subject to the limitations of paragraph 2 of Article 10. This is consistent with the ECtHR's broader approach, which is permissive of content and viewpoint-based proscriptions on expression for the protection of the reputation and rights of others.

To date, the Court has addressed state efforts to regulate disinformation only once, in its 2019 ruling in *Brzeziński v Poland*.[76] This case concerned a provision in Poland's election law that allowed a court, within 24 hours, to consider whether 'untrue information' had been published, and to issue an order prohibiting its further distribution. Reiterating that there was little scope under Article 10 for restrictions on political speech, the Court emphasised that the lower courts did not appear to have examined whether the impugned expression had a credible factual basis

[72] See *Pastörs* (n 68) para 47.
[73] ibid.
[74] ibid para 48.
[75] ibid. See also *Erbakan v Turkey*, App no 59405/00 (ECtHR, 6 July 2006) (in which the Court considered the case of a Turkish politician and prior Prime Minister, who, while serving as a chairman of one of the country's political parties, gave a speech during a local election campaign rally in which he made comments that later resulted in a criminal conviction for incitement to hatred or hostility on the basis of religion and race. In its holding, the Court emphasised that combatting all forms of intolerance was an integral component of the protection of human rights and that it was important that politicians should avoid making speeches with comments likely to foster such intolerance).
[76] App no 47542/07 (ECtHR, 25 July 2019).

or whether the applicant had acted with the requisite diligence.[77] Rather, the impugned speech had been immediately characterised as false and damaging to the reputation and standing of the complainant. Accordingly, the Court concluded that a fair balance was not struck between the applicant's Article 10 rights and the need to protect the complainants' rights and reputation.[78] While *Brzeziński* is notable for addressing disinformation, it is not particularly instructive as the scrutiny applied by the lower court was clearly inadequate and lacking in rigour.

Overall, a review of the ECtHR's jurisprudence in this area reveals a lack of clarity in its approach to Article 10 challenges involving state actors, in particular, with regard to false statements of fact. On the one hand, the Court recognises that persons in positions of influence, including elected officials, have a responsibility to speak on matters of public concern in ways that do not undermine important democratic principles such dignity.[79] On the other hand, the Court emphasises that such actors, in particular legislators, are deserving of a *higher* level of speech protection than 'ordinary' citizens on account of their unique status and function in the democratic order.[80] This lack of clarity in the ECtHR's jurisprudence, coupled with the recent report from the Committee on Human Rights, highlight the inherent challenges in delineating the scope of protection of the expression of state actors and the ways in which courts grapple with their unique status within frameworks that provide expansive protection to political expression whilst also attaching duties and responsibilities to the exercise of rights, which balance the rights of the speaker against the rights of the targets of extreme speech.

B. Regulating the Speech of State Actors in the United States: A Tangled Web of Private, Official, and Government Speech

The American immunity framework is comprised of constitutional, legislative, and doctrinal elements, which provide for a limited set of immunities to state actors when engaging in speech within the scope of their official duties. The Speech and Debate Clause of Article I of the Constitution instructs that 'any Speech or Debate in either House, [Senators and Representatives] shall not be questioned in any other Place'.[81] It protects Congressmembers

[77] 'Breach of the right to freedom of expression in the context of an electoral campaign (Press Release)' (*Registrar of the ECtHR*, 2019) ECHR 284.
[78] ibid.
[79] See, e.g., *Willem* (n 65).
[80] See, e.g., *Castells* (n 69).
[81] US Const art I, §, 6, cl 1. See also *Kilbourn v Thompson*, 103 US 168 (1880).

against prosecutions that directly impinge upon or threaten the legislative process.[82] However, the immunity is limited in that it applies to speech made in the context of 'legislative activities' or the 'legislative process', and does not extend beyond what is necessary to preserve the integrity thereof.[83] As discussed below, both the legislative and judicial branches have extended principles of immunity well beyond the scope of Article I.

1. The Statutory Framework for Immunity

The Federal Tort Claims Act (FTCA)[84], as amended by the Westfall Act[85], provides that '[t]he United States, as sovereign, is immune from suit save as it consents to be sued'.[86] While the FTCA waives sovereign immunity for certain torts committed by federal employees while acting in the scope of their employment, it does *not* waive such immunity for claims of state actors arising out of libel and slander.[87] For the purposes of the Westfall Act, a determination of whether an employee was acting within the scope of employment is a question of law made in accordance with the law of the state where the conduct occurred.[88] While the USSC has not ruled on the question of whether Members of Congress are government employees for purposes of the Westfall Act, the Federal Circuit Courts that have opined on the issue ruled that these actors are employees or officers of the federal government.[89]

A recent application of the Westfall Act in relation to online speech from Members of Congress is instructive. In 2020, the Sixth Circuit Court of Appeals considered the question of whether allegedly defamatory tweets sent by Representative Debra Haaland and Senator Elizabeth Warren, in response to a widely publicized incident involving an encounter between Trump supporters and a Native American activist, were made in the scope of employment under Kentucky law for the purposes of the Westfall Act.[90] The court held that the tweets were made

[82] *Gravel v United States*, 408 US 606, 616 (1972).
[83] *Hutchinson v Proxmire*, 443 US 111, 127 (1979).
[84] 28 USC ss 1346, 2671 *et seq.*
[85] ibid s 2679(b)(1).
[86] See *United States v Sherwood*, 312 US 584, 586 (1941) (internal citations omitted).
[87] FTCA (n 84) 1346, 2671 *et seq*. The term 'employee of the government' includes officers or employees of any federal agency. The term 'federal agency' includes the executive departments, the judicial and legislative branches, and other governmental entities.
[88] *Dolan v United States*, 514 F3d 587, 593 (6th Cir 2008).
[89] See, e.g., *Williams v United States,* 71 F3d 502, 504 (5th Cir 1995); *Operation Rescue Nat'l v United States*, 147 F3d 68, 70 - 71 (1st Cir 1998).
[90] *Does v Haaland*, 973 F3d 591 (6th Cir 2020). The court reasoned that the allegedly defamatory tweets occurred in Kentucky 'because Plaintiffs live in Kentucky and the tweets were accessible in that state'. ibid 599. Senator Warren's impugned speech included a tweet from her official Senate Twitter account that stated 'Omaha elder and Vietnam War veteran Nathan Phillips endured hateful taunts with dignity and strength, then urged us all to

in the scope of the legislators' employment, finding no meaningful difference between tweets and the other kinds of public communications between elected officials and their constituents held to be within the scope of employment under the statute. The court reasoned that tweeting fits within the 'wide range of legitimate "errands" performed for constituents,' which include 'preparing so-called "newsletters" to constituents, news releases, and speeches delivered outside the Congress'.[91] It characterised the statements, which were critical of particular Trump supporters, as 'calculated to serve the interests of Congressmembers' constituents by informing them of Congressmembers' views regarding a topical issue'.[92]

A recent filing by the DOJ in a case pending in the Second Circuit Court of Appeals, involving allegedly defamatory statements made in 2020 by then President Trump, provides insight into the American government's position on the scope of immunity for the expression of federal employees.[93] The case of *Carroll v. Trump* concerns whether and to what extent the FTCA and the Westfall Act apply to the office of the president. The facts concern allegedly defamatory statements made to the press by then President Trump in denying E. Jane Carroll's publicly disclosed allegations that he had raped her two decades earlier.[94] Following these statements, Ms. Carroll sued then President Trump for defamation in his individual capacity in state court.[95] For nearly a year the lawsuit proceeded as an ordinary defamation case between two private parties. Following the state court's denial of then President Trump's claim that he could not be sued while serving as president, the United States government removed the case from state to federal court based on the argument that the statements to the press regarding Ms.

---

do better. Listen to his words'. Representative Haaland's impugned speech included a tweet from her official Congressional Twitter account that read: 'This Veteran [Nathan Phillips] put his life on the line for our country. The students' display of blatant hate, disrespect, and intolerance is a signal of how common decency has decayed under this administration. Heartbreaking.' ibid 594.

[91] ibid 602 (citing *United States v Brewster*, 408 US 501 (1972)).

[92] FTCA (n 84) s 2679(b)(1). *See also Williams v United States*, 71 F3d 502, 505 (5th Cir 1995) (holding that a Congressman's allegedly defamatory statements about a lobbyist made during a press interview about Congress' appropriation of certain federal monies were within the scope of his employment and he was therefore immune from a defamation suit under the Westfall Act because the statements, including the allegedly defamatory ones, were made in performance of his duty to 'inform constituents and the public at large of the issues being considered by Congress'); *Operation Rescue Nat'l v United States*, 975 FSupp92, 106 (D Mass 1997) (holding that then Senator Ted Kennedy was immune from liability based on statements made during the course of responding to a reporter's question at Kennedy's home pertaining to a bill he was sponsoring that addressed access to women's health clinics in which he said that anti-abortion organizations like the plaintiff had a 'national policy [of] firebombing and even murder' because he was 'was providing political leadership and a basis for voters to judge his performance in office and, in that sense his 'employer was his constituents and he served them by fully informing them of his views and working to pass legislation he believed would benefit them.').

[93] *Carroll v Trump,* Case 20-2977, Doc 102 (2nd Cir 7 June 2021).

[94] *Carroll v Trump*, 498 FSupp3d 422 (SD NY 2020). These statements included then President Trump calling Ms Carroll a liar, claiming that she had made up the rape allegations, and that he had never met her.

[95] ibid.

Carroll were within the scope of Mr. Trump's duties as president. Additionally, the government moved to substitute the United States as the defendant in the case based on the argument that Ms. Carroll sued an 'employee' of the United States for actions within the scope of his employment under the FCTA and the Westfall Act.

The lower court held that the president is not an officer of a federal agency under the Westfall Act and, thus, there was no statutory basis to trigger the Attorney General's obligation to defend a government employee - with the government the substituted defendant under the FTCA - and that the allegedly defamatory statements were not made within the scope of Mr. Trump's position as president. On appeal to the Second Circuit, the DOJ is challenging the lower court's resolution of both issues. In arguing that former President Trump's statements regarding Ms. Carroll fall under the Westfall Act, the DOJ argues that 'speaking to the public and the press on matters of public concern is undoubtedly part of an elected official's job', even when such matters relate exclusively to the private life of the official.[96] The DOJ refutes the lower court's contrary conclusion that rested on the fact that then President Trump's views on Ms. Carroll's rape allegations were not pertinent to his employment because 'they reveal[ed] nothing about the operation of government' and '[n]either the media reports nor the underlying allegations have any relationship to his official duties'.[97] The DOJ also relies on case law from a lower court suggesting that the 'scope of employment' under the Westfall Act may cover conduct involving 'serious criminality' and/or conduct that 'runs contrary to the national security of the United States'.[98] The DOJ further argues that the United States should be responsible for the intentional torts of its employees even if prompted, in part, by personal motives.[99]

2. The Doctrinal Framework for Immunity

During the twentieth century, the USSC developed an extensive framework for immunities for the speech of state actors. With respect to legislators, the USSC holds that the 'manifest function of the First Amendment in a representative government requires that legislators be given the widest latitude to express their views of policy'.[100] More broadly, the Court holds that the role elected officials play in society 'makes it all the more imperative that they be

[96] *Carroll* 'Doc 102' (n 93) 3.
[97] ibid 15.
[98] ibid. See also *Rasul v Myers*, 512 F3d 644, 660 (DC Cir 2008) and *Wilson v Libby*, 535 F3d 697, 711 - 12 (DC Cir 2008).
[99] ibid (*Carroll*). See also *Jacobs v Vrobel*, 724 F3d 217 (DC Cir 2013); *District of Columbia v Bamidele*, 103 F3d 516, 525 (DC Cir 2014).
[100] *Bond v Floyd*, 385 US 116, 135 - 36 (1966).

allowed freely to express themselves on matters of current public importance'.[101] The Court recognises an absolute privilege for the expression of federal officials if made 'within the outer perimeter' of their duties.[102] Unlike the ECtHR, the USSC does not attach duties or responsibilities to the speech of state actors or politicians and, as discussed in Chapter 2, rejects the principle that those exercising their right to freedom of expression undertake any duties or responsibilities relating thereto.[103]

The USSC's jurisprudence concerning expression from state actors identifies three general categories of speech: private, official and government. The distinction between a state actor speaking in a private rather than professional capacity was considered by the Court in *Wood v. Georgia*, which involved an elected sheriff who argued that the Georgia courts violated his First Amendment rights by holding him in contempt of court for expressing opinions on a matter before a grand jury in his county. The Court rejected the state's argument that the petitioner's status as sheriff meant that he owed a 'special duty and responsibility to the court and its judges' and that this right to freedom of expression should be 'more severely curtailed than the average citizen'.[104] Of particular relevance to the Court was that the prosecution did not rely on the petitioner's status to show a more substantial likelihood that his statements would disrupt the administration of justice, and that they were not intended to hinder the investigation.[105] For these reasons, the Court concluded that the plaintiff spoke in his personal rather than professional capacity and highlighted the dearth of evidence suggesting that the statements interfered with the sheriff's performance of his 'official' duties.[106] In the absence of some other showing of a 'substantive evil actually designed to impede the course of justice in justification of the exercise of the contempt power to silence the petitioner', the Court held the statements were protected speech.[107]

---

[101] *Wood v Georgia,* 370 US 375, 395 (1962).
[102] *Barr v Matteo*, 360 US 564, 575 (1959).
[103] See Chapter 2, pp 77 - 79.
[104] *Wood* (n 101) 393.
[105] ibid 382.
[106] ibid 384. Specifically, the Court observed that it was 'not dealing with a situation where a sheriff refuses to issue summonses or to maintain order in the court building; nor, so far as the record shows, did the petitioner do any act which might present a substantive harm to the jury's solution of the problem placed before it. We are dealing here only with public expression'. ibid.
[107] ibid. See also *Bond* (n 100) (holding that the Georgia Legislature violated the First Amendment when it refused to seat an elected state legislator because of his comments criticising America's involvement in the Vietnam War). As in *Wood* (n 101), the Court reasoned that the state legislator's expressive activity could not be censored simply because of the nature of his position. It is important to note that in each of these cases, the Court's holding hinged on the fact that the impugned statements were made in the speaker's role as a private citizen, rather than a public official.

The distinction between private, official, and government speech is significant because each category of expression raises distinct First Amendment issues. As in *Wood*, state actors speaking as private citizens trigger the protections afforded under the First Amendment with regard to that speech. State actors speaking in their official capacities trigger the application of the state action and public forum doctrines, which subject the speech to the constraints of the First Amendment. As discussed in Chapter 1, the state action doctrine draws the line between governmental and private conduct (with only the former implicating free speech protections) and the public forum doctrine instructs courts as to what level of protection from government interference contested speech warrants in a particular case.[108]

Finally, if speech is designated as 'government speech', the First Amendment is not implicated at all. This is due to the USSC's government speech doctrine, which holds that '[t]he Free Speech Clause...does not regulate government speech'.[109] Put another way, when the 'government speaks' it is not barred by the Free Speech Clause from determining the content of what it says.[110] The First Amendment is thus interpreted as not requiring that Congress and other government entities abridge their own ability to speak freely. The reasoning underlying the doctrine is that imposing a requirement of viewpoint neutrality on government speech would be 'paralysing', as it must take particular viewpoints and reject others when it engages in the process of governing.[111] Accordingly, the government is entitled to promote a program, to espouse a policy, or to take a position.[112] For example, the government may support valid programs and policies by taxes or other exactions binding on protesting parties without triggering the First Amendment, and such funds may be spent for speech to advocate or defend such programs and policies.[113]

---

[108] See Chapter 1, pp 46 - 52.
[109] *Matal v Tam,* 137 SCt 1744, 1758 (2017).
[110] *Pleasant Grove City v Summum*, 555 US 460, 467 - 469 (2009).
[111] See, e.g., *Matal* (n 109) 1758 (in which the Court opined that '[d]uring the Second World War, the Federal Government produced and distributed millions of posters to promote the war effort. These posters expressed a viewpoint, but the First Amendment did not demand that the Government balance the message of these posters by producing and distributing posters encouraging Americans to refrain from engaging in these activities.'
[112] *Walker v Texas Div, Sons of Confederate Veterans, Inc*, 576 US 200 (2015). See also *Johanns v Livestock Marketing Assn*, 544 US 550, 574 - 575 (2005): '[t]o govern, government has to say something, and a First Amendment heckler's veto of any forced contribution to raising the government's voice in the "marketplace of ideas" would be out of the question.' (internal citations and quotations omitted).
[113] *Johanns* (ibid) 559.

One of the rationales underpinning the government speech doctrine is that government statements do not normally trigger the rules designed to protect the marketplace of ideas.[114] The freedom for the government to determine what it says without the constraints of the Free Speech Clause is premised on the notion that it is the democratic electoral process that first and foremost provides a check on government speech. That is, the Free Speech Clause helps produce *informed* opinions among members of the public, who then 'influence the choices of a government that, through words and deeds, will reflect its electoral mandate'.[115] This is tied to the notion that the political process serves as a check on what the government chooses to say because if enough voters object to government speech, 'the next election effectively cancels the message'. [116]

The government speech doctrine is both relatively new and relatively imprecise. While not explicitly mentioned in *Rust v Sullivan,*[117] this is generally considered the first case in which the USSC applied the doctrine.[118] While the USSC cautions that the doctrine is susceptible to dangerous misuse, it provides little insight regarding what such misuse might look like and the proper mechanisms by which to address it.[119] In *Matal,* the USSC provided some clarity by opining 'that if private speech could be passed off as government speech by simply affixing a government seal of approval, government could silence or muffle the expression of disfavoured viewpoints'.[120]

Within this ambiguity and uncertainty, the digital age has brought new challenges with respect to distinguishing between private, official, and government speech. For example, federal courts across the country have recently adjudicated disputes over the correct designation of speech appearing on legislator's official social media accounts. These distinctions are significant because, as explained above, the way in which speech is classified impacts its

[114] *Walker* (n 112) 200. For an overview of the marketplace of ideas, see Chapter 3, p 112 - 114.
[115] *Stromberg v California*, 283 US 359 (1931). In *Stromberg,* the Court further opined that the American constitutional system seeks to maintain 'the opportunity for free political discussion to the end that government may be responsive to the will of the people'. ibid 369.
[116] See, e.g., *Walker* (n 112) 200: 'it is the democratic electoral process that first and foremost provides a check on government speech'.
[117] 500 US 173 (1991).
[118] See e.g., *Foxworthy v Buetow*, 492 FSupp2d 974, 984 (SD Ind 2007): '[i]n *Rust v Sullivan*…the Supreme Court first applied the government speech doctrine, albeit not naming it as such'.
[119] The Court seems mostly concerned with distinguishing private speech from government speech and also states, without any detail, that the involvement of public officials in advocacy may be limited by law, regulation, or practice. See *Pleasant Grove* (n 110) 468.
[120] *Matal* (n 109) 1757.

treatment under the First Amendment. For example, in *Faison v. Jones,*[121] the Eastern District of California designated a county sheriff's *posts* on his Facebook page as government speech that fell outside the First Amendment's constraints, but the *administration* of such accounts, including decisions to block users, as official speech, i.e., state action on a public forum, which precluded the official from blocking users based on the content or viewpoint expressed in their posts.[122]

An examination of the treatment of state actors in European and American free speech frameworks reveals intersecting and, at times, inconsistent principles, highlighting the challenges of regulating this type of expression within broader frameworks primarily concerned with the dangers posed by the regulation of private speech. The key challenges for the ECtHR involve reconciling the tension between parliamentary immunity and the recognition that certain types of speech undermine democratic values and are therefore unworthy of protection, and that all individuals - even state actors - undertake duties and responsibilities when speaking in public discourse, and perhaps greater duties and responsibilities than 'regular' citizens. This serves to place limitations on the views that state actors may express, even within the sphere of parliamentary immunity.

There is no such tension in the American framework because considerations of whether particular speech undermines democracy are irrelevant to the protection afforded to speech in public discourse and no speaker, regardless of their status, undertakes duties and responsibilities in relation to the exercise of the right to freedom of expression. The inherent challenges, instead, relate to the correct classifications for speech from state actors, which fundamentally impact the treatment of such speech for First Amendment purposes. Moreover, the judicial, legislative, and the executive branches have adopted a sweeping interpretation of immunity in this context, applying an approach that considers any statements, however tangentially related to the duties of state actors, as within the scope of their employment so as to trigger relevant immunities. Additionally, unlike in Europe, the influence that state actors wield in public discourse on account of their status is largely disregarded by American courts. Instead, the government speech doctrine functions to classify certain types of speech from state

[121] 440 FSupp3d 1123 (ED Cal 2020).
[122] See also *Leuthy v Lepage*, 2018 WL 4134628 (D Maine 2018); *Davison v Randall*, 912 F3d 666, 681 (4th Cir 2019); *Garnier v Poway Unified Sch Dist*, 2019 WL 4736208 (SD Cal 2019).

actors as entirely outside the scope of the First Amendment. Chapter 6 explores the extent to which immunity doctrines and the government speech doctrine remain fit for purpose in the digital age.

IV. Disinformation from State Actors in the Digital Age: A New and Dangerous Chapter of Lying in Politics

In its Freedom on the Net 2019 report, Freedom House emphasised that governments around the world are increasingly using social media to manipulate elections and monitor their citizens, and that disinformation is, by far, the most popular tactic for digital election interference.[123] Among Freedom House's findings is that domestic actors interfered in twenty-six out of the thirty countries that held elections or referendums in 2018 and during this period in the United States false, misleading, and/or hyper-partisan online content proliferated.[124] The Oxford Disinformation Report recorded similar findings, including that seventy-six countries used disinformation and media manipulation to mislead social media users.[125]

The rise of social media has increased the use of disinformation as a campaigning tactic with an impact 'far beyond the first iteration'. [126] Recent research suggests that online, lies 'travel significantly farther, faster, deeper, and more broadly than the truth', particularly in the context of political news.[127] Social media also enables purveyors of disinformation to target messages with more precision than is possible with traditional media.[128] Another advantage of social media to such purveyors is that disinformation is subject to less scrutiny, as the traditional media is unable to effectively perform its gatekeeping function, including filtering information

[123] Allie Funk and Adrian Shabhaz, 'Freedom on the Net 2019: The Crisis of Social Media' (*Freedom House,* 2019) <https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf > accessed 10 February 2020. According to Freedom House's President, '[a]uthoritarians and populists around the globe are exploiting both human nature and computer algorithms to conquer the ballot box, running roughshod over rules designed to ensure free and fair elections.'. ibid 1.
[124] ibid.
[125] Bailey (n 8) 15.
[126] Caroline Fisher and Ivor Gaber, '"Strategic Lying": The Case of Brexit and the 2019 U.S. Election' (2021) Int J of Press/Pol <doi:10.1177/1940161221994100> 5.
[127] Sinan Aral, Deb Roy and Soroush Vosoughi, 'The Spread of True and False News Online' (2018) 359(6380) Science 1146. In this study, researchers at the Media Lab at the Massachusetts Institute of Technology, investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprise ~126,000 stories tweeted by ~3 million people more than 4.5 million times. News was classified as true or false using information from six independent fact-checking organisations that exhibited 95 to 98 percent agreement on the classifications.
[128] Fisher and Gaber (n 126) 5.

through professional editing and fact checking.[129] Moreover, due to the sheer volume of information on social media it is impracticable for the traditional media to verify every questionable tweet or post.[130] State actors in Europe and the United States are leveraging the advantages of online disinformation to wage campaigns aimed at improperly manipulating public opinion. Section A examines whether and to what extent recent disinformation campaigns from domestic state actors endanger democratic norms and institutions.

A. Disinformation in Europe: A Growing But (So Far) Manageable Challenge to the EU and European Democracies

This Section primarily focuses on the disinformation in the EU and EU Member States. The European Commission describes large-scale disinformation campaigns as 'a major challenge for Europe', requiring a coordinated response from EU countries, EU institutions, social networks, traditional media, and EU citizens.[131] Perhaps the worst EU Member State in terms of domestic disinformation is Hungary; the Hungarian government has been running ongoing disinformation campaigns since 2019 in addition to '[financing] an entire fake news industry'.[132] These campaigns involve spreading far-right stereotypes on every distribution channel available to the government, including social media.[133] Following an anti-migrant campaign, which started in 2015 and reinforced the 'ethno-nationalist boundaries of Hungarian-ness', the Hungarian population was found to be more xenophobic than at any time in the past 25 years.[134] The campaign arguably served to normalise and mainstream extreme-right beliefs and language, moving extreme-right ideology and rhetoric that are acceptable in public discourse further to the right.[135]

[129] ibid 5 - 6.
[130] ibid 6.
[131] European Commission (n 12).
[132] Samantha Bradshaw and Philip N Howard, 'Online Supplement to Working Paper 2018.1 Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation' (*Oxford Internet Institute*, 2019) 29 <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct_appendix.pdf> accessed 6 June 2021.
[133] Bayer (n 10) 44 - 45.
[134] ibid. See also Christian Keszthelyi, 'Xenophobia Skyrocketing in Hungary, Surveys Reveal' (*Budapest Business Journal,* 17 November 2016) <https://bbj.hu/budapest/xenophobia-skyrocketing-in-hungary-surveys-reveal_124920> accessed 3 November 2020, finding that that 'xenophobia saw a hike in Hungary when the refugees had disappeared and the campaign against "migrants" by the Hungarian government accelerated'.
[135] ibid.

The problem of domestic disinformation from state actors in the EU is not limited to weakened democracies.[136] For example, a 2020 report compiled by the Oxford Internet Institute found that computational propaganda is a widespread tactic amongst multiple actors in the British political system.[137] The study highlighted the Coalition for Reform in Political Advertising's description of political advertising coming from the main parties during the General Election in December 2019 as 'illegal, indecent, dishonest and untruthful', with advertising that 'transgressed' deriving from the Conservative, Labour and Brexit Parties, as well as the Liberal Democrats.[138] It further notes that analysts reported on the 'apparent impunity with which the main parties...employed overt disinformation to secure votes'.[139]

In response to the dangers to democratic norms and institutions flowing from online disinformation, the European Commission has developed a number of initiatives, including the EU wide voluntary Code of Practice on Disinformation (Code of Practice), in force since 2018, which lays out a set of worldwide self-regulatory standards for industry with the objective of limiting the spread and impact of disinformation,[140] the European Digital Media Observatory, which consists of a European hub for fact-checkers, academics and other relevant stakeholders to support policy-makers, the Action Plan on Disinformation, which aims to strengthen EU capability and cooperation in the fight against disinformation,[141] and the European Democracy Action Plan, which aims to develop guidelines for obligations and accountability of online platforms.[142] Additionally, the EU has undertaken and commissioned several studies on the impact of online disinformation in the region.[143]

---

[136] According to Freedom House, Hungary's democratic decline 'has been the most precipitous ever tracked in *Nations in Trust*; it was one of the three democratic frontrunners as of 2005, but in 2020 it became the first country to descend by two regime categories and leave the group of democracies entirely'. Zselyke Csaky, 'Nations in Transit 2020: Dropping the Democratic Façade in Europe and Eurasia' (*Freedom House,* 2020) 5 <https://freedomhouse.org/report/nations-transit/2020/dropping-democratic-facade> accessed 13 February 2021.
[137] Samantha Bradshaw and others, 'Country Case Studies Industrialised Disinformation: 2020 Global Inventory of Organised Social Media Manipulation' (*Oxford Internet Institute*, 2020) <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/03/Case-Studies_FINAL.pdf> accessed 3 March 2021, 415
[138] ibid 415.
[139] ibid. See also Jente Althuis, Francesca Granelli and Thomas Colley, 'Disinformation's Societal Impact: Britain, Covid, And Beyond' (2020) 8 Defence Strategic Communications 89.
[140] The Code's signatories include Facebook, Google, Twitter, Mozilla, which provide monthly reports to the European Commission.
[141] The Action Plan lays out policy norms to restrict disinformation.
[142] European Commission (n 12).
[143] See, e.g., Alexandre Alaphilippe and others, 'Automated tackling of disinformation' (*Publications Office of the European Union*, 2019) <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU(2019)624278_EN.pdf> accessed 21 May 2021; Stephan Lewandowsky and others, 'Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making (EU Joint Research Centre)' (*Publications Office of the European Union*, 2020) doi:10.2760/709177.

In May of 2021, the European Commission issued Guidance on Strengthening the Code of Practice, which calls for reinforcing the Code of Practice in certain areas to ensure a complete and consistent application across stakeholders and EU countries, including creating a more robust monitoring framework and providing increased access to data to researchers.[144] This Guidance emphasises that the EU approach to countering disinformation is grounded in the protection of freedom of expression and that, as a result, rather than criminalising or prohibiting disinformation, the EU strategy aims to make the online environment and its actors more transparent and accountable.[145]

Efforts to combat disinformation are also being undertaken at the national level in Europe. France is at the forefront of national efforts to stem the tide of disinformation by way of compulsory regulation of intermediary platforms. In 2018, the French parliament passed two laws directed to banning 'fake news', primarily during election cycles, on the basis that disinformation may harm the functioning of democratic institutions.[146] The laws target the dissemination of fake news by means of digital tools, in particular, those used on social media.[147] In order to stop 'the rapid spread of fake news', conclusions regarding 'whether fake news is manifest and disseminated deliberately on a massive scale, and whether this had to a disturbance of the peace or has influenced the results of an election, will be referred to an interim judge'.[148] The laws also confer new responsibilities upon the French media regulator, the Conseil Supérieur de l'Audiovisuel (CSA), which has the authority to act against any media distributing disinformation. Once the CSA has notified a company regarding the intentional transmittal of false information, the company has only 48 hours to argue for continued transmission.[149] The legislation also establishes a duty of cooperation for digital intermediaries, forcing them to introduce, and make publicly available, measures to eliminate fake news.[150]

---

[144] 'European Guidance on Strengthening the Code of Practice on Disinformation ' (Communication) COM (2021) 262 final.
[145] ibid.
[146] LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information.
[147] Gouvernement de France, 'Against information manipulation' <https://www.gouvernement.fr/en/against-information-manipulation> accessed 1 July 2021.
[148] Gouvernement de France, 'A bill against the manipulation of information' (7 June 2018) <https://www.gouvernement.fr/en/a-bill-against-the-manipulation-of-information> accessed 1 July 2021.
[149] See Rachael Craufurd Smith, 'Fake news, French law and democracy: lessons for the United Kingdom?' (2019) 11(1) J of Med L 52, 59.
[150] Gouvernement de France (n 147).

France's efforts to tackle disinformation by way of legislation has been subject to criticism for threatening freedom of speech and limiting democratic debate.[151] For example, in practice, the judge might not be able to make a decision regarding the veracity of content in just 48 hours. Moreover, given how quickly disinformation spreads on social media, the efficacy of such an approach is rightly questioned. France's foray into compulsory regulation of intermediary platforms with respect to controlling the spread of online disinformation highlights the broader challenges of government efforts to regulate private expression on social media platforms. Disinformation laws that are too broad and vague, like those discussed in relation to 'hate speech' and terrorist-related expression, present a significant risk of chilling legitimate speech, and may be used selectively and/or indiscriminately to encourage or force private companies to moderate their platforms in ways that can harm freedom of expression and stifle public discourse.[152] These concerns are discussed in detail in Chapter 6.

B. Domestic Disinformation in the United States: An Epistemic Crisis that Poses a Direct Threat to American Democracy

The proliferation of disinformation infecting American public discourse has recently been described as an epistemic crisis.[153] While disinformation from foreign actors remains a problem in the United States, including recent pandemic disinformation disseminated by both China and Russia, domestic disinformation is a much larger and more serious issue.[154] For example, a 2018 study by researchers at Oxford University found that the majority of disinformation concerning the 2018 midterm elections on Twitter and Facebook came from domestic sources.[155] The most instructive example of the challenges to American democracy posed by domestic disinformation from state actors concerns the campaign led by former

[151] ibid.
[152] MacKenzie F Common and Rasmus Kleis Nielsen, 'Submission to UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on disinformation' (*Reuters Institute*, 15 February 2021) <https://reutersinstitute.politics.ox.ac.uk/risj-review/how-respond-disinformation-while-protecting-free-speech> accessed 5 July 2021.
[153] See Yochai Benkler, Robert Faris and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (OUP 2018).
[154] See Summer Lopez and James Tager, 'Truth on the Ballot: Fraudulent News, the Midterm Elections, and Prospects for 2020' (*Pen America,* 13 March 2019) <https://pen.org/wp-content/uploads/2019/03/Truth-on-the-Ballot-report.pdf> accessed 20 June 2021.
[155] Nahema Marchel and others, 'Polarization, Partisanship and Junk News Consumption on Social Media During the 2018 US Midterm Elections' (*Oxford Internet Institute,* 1 November 2018) <https://demtech.oii.ox.ac.uk/research/posts/polarization-partisanship-and-junk-news-consumption-on-social-media-during-the-2018-us-midterm-elections/> accessed 28 January 2021.

President Trump and his allies to undermine the public's faith in the integrity of the 2020 election. This section uses this disinformation campaign as a case study of the harms caused by online disinformation from state actors in the digital age.

By all measures, the 2020 general election was one of the most secure elections in American history with no evidence of widespread voter fraud.[156] These facts were confirmed over and over again by federal agencies, including the FBI, the DHS, and the DOJ, as well as the United States Election Assistance Commission and dozens of federal courts.[157] President Biden won the Electoral College handily - 306 to 232 - and it is worth noting won the popular vote by over 7 million votes.[158] In the past six presidential elections, only Barack Obama in 2008 won by a larger total vote margin. Additionally, with the exception of Former President Obama's victory in 2008, President Biden's 4.5 percentage point lead in the popular vote represented the largest victory in the past six presidential elections.[159] Finally, voter turnout for the 2020 election was the highest in 120 years in terms of the percentage of the voting-eligible population.[160]

Notwithstanding this fact, former President Trump, with what appeared to be the coordinated support of the Republican National Committee (RNC) and members of his re-election campaign staff, waged a six-month institutionalised disinformation campaign aided by the right-wing media ecosystem for the purpose of undermining the public's faith in the integrity of the 2020 election.[161] A Harvard study (Harvard Disinformation Study) released just prior to the November 2020 election analysed fifty-five thousand online media stories, five million tweets, and seventy-five thousand posts on public Facebook pages that referred to mail-in

[156] Following the election, the DHS's Cybersecurity & Infrastructure Agency released a statement from the coordinating bodies on American election infrastructure and security that stated that it was 'the most secure in American history' and that 'there [wa]s no evidence that any voting system deleted or lost votes, changed votes, or was in any way compromised.' See 'Joint Statement From Elections Infrastructure Government Coordinating Council & the Election Infrastructure Sector Coordinating Executive Committees' (*CISA*, 12 November 2020) <https://www.cisa.gov/news/2020/11/12/joint-statement-elections-infrastructure-government-coordinating-council-election> accessed 28 January 2021.

[157] For a comprehensive overview of the evidence of the integrity of the 2020 election, see 'It's Official: The Election Was Secure' (*Brennan Center for Justice,* 11 December 2020) <https://www.brennancenter.org/our-work/research-reports/its-official-election-was-secure> accessed 4 February 2021.

[158] See James M Linsday, 'The 2020 Election by the Numbers' (*Council on Foreign Relations,* 15 December 2020) <https://www.cfr.org/blog/2020-election-numbers> accessed 14 February 2020.

[159] ibid.

[160] ibid.

[161] See Yochai Benkler and others, 'Mail-In Voter Fraud: Anatomy of a Disinformation Campaign' (2020) Berkman Center Research Publication No 2020 <https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020> accessed 5 July 2021.

voting and the risk of fraud, all posted between March 1 and August 31 of 2020.[162] It notes that the RNC and re-election campaign staff were 'repeatedly and consistently on message at the same moments' as then-President Trump's tweets, 'suggesting an institutionalised rather than individual disinformation campaign' that was supported by 'the right-wing media, primarily Fox News and talk radio functioning as a party press' that 'reinforced the message, provide[d] the president a platform, and marganlize[d] those Republican leaders or any conservative media personalities who insist[ed] that there [wa]s no evidence of widespread voter fraud associated with mail-in voting'.[163]

1. 'You're the President, not Someone's Crazy Uncle':[164] Why the Status of the Speaker Matters

One of the most significant conclusions from the Harvard Disinformation Study is that contrary to most contemporary analyses of disinformation efforts in the American information ecosystem, the disinformation campaign involving mail-in voter fraud, which manipulated the views of tens of millions of American voters, did not originate in social media or via Russia or some other foreign adversary. Rather, it was led by former President Trump, fomented by prominent members of the Republican Party, and amplified by some of the biggest media outlets in the country, with social media playing a secondary and supportive role.[165] These findings are consistent with a separate study by the same researchers from 2015 - 2018, which found that Fox News and Mr. Trump's campaign were, collectively, far more influential in spreading false beliefs than Russian trolls or Facebook 'clickbait artists'.[166] The researchers posited that this dynamic was likely even more pronounced in the 2020 election cycle because of Mr. Trump's position as president and his role as head of the Republican Party, which allowed him to 'operate directly through political and media elites, rather than relying exclusively on online media as he did when he sought to advance his then-still-insurgent positions in 2015 and the first half of 2016'.[167]

---

[162] ibid. See also Center for an Informed Public, Digital Forensic Lab, Graphika, and Stanford Observatory, 'The Long Fuse: Misinformation and the 2020 Election' (*Stanford Digital Repository*, 2021) <https://purl.stanford.edu/tr171zs0069> accessed 5 July 2021.
[163] Benkler 'Mail-In Voter Fraud' (n 161) 2.
[164] NBC News host Savannah Guthrie interviewing then President Trump in October of 2020 regarding his decision to retweet a QAnon conspiracy theory involving Joe Biden to his 87 million Twitter followers.
[165] Benkler 'Mail-In Voter Fraud' (n 161) 2.
[166] See Benkler 'Network Propoganda' (n 153).
[167] Benkler 'Mail-In Voter Fraud' (n 161) 1.

Ultimately, the Harvard Disinformation Study concluded that the 'usual suspects' in public debates about disinformation, including fake pages created by foreign adversaries or Facebook's algorithms, did not explain any peak engagement that was not *better* explained as having been initiated and heavily fomented by political figures and elite right-wing media personalities, and disseminated to millions of American by major media outlets. Additionally, the role played by bots or trolls on Twitter was dwarfed by tweets from then President Trump, his staff, and other institutional and media allies, including the RNC and Fox News.[168]

Polling conducted before and after the presidential election suggest that this disinformation campaign was hugely successful. Polling conducted in September of 2020 revealed that nearly half of Republicans believed then President Trump's lie that election fraud was a major concern associated with expanded mail-in voting during the pandemic and viewed Democrats as the most likely perpetrators of election interference.[169] Polling conducted after the Capitol insurrection found that nearly seventy-two percent of likely Republicans questioned the results of the presidential election, many doing so on the basis of specious allegations of voter fraud.[170] In a Monmouth University Poll conducted between 25 February and 1 March 2021, approximately one-third of Americans reported believing that President Biden's victory was the result of voter fraud.[171] Among Republicans, nearly sixty-five percent believe President Biden's victory was due to voter fraud and twenty-nine percent reported that 'they will never accept him as President'.[172] A Reuters/Ipsos poll conducted in May 2021 revealed similar attitudes with over half of Republicans believing Donald Trump is the actual President of the United States and that the election was 'rigged' or the result of voter fraud.[173] These attitudes

---

[168] ibid 6 - 7.
[169] ibid 1.
[170] For example, a Quinnipiac University poll released on 11 January found that seventy-three percent of Republicans believed in widespread voter fraud, compared to five percent of Democrats. See Data for Progress and Vox, 'Poll on Election Trust' (*Quinnipiac University*, 2021) <https://www.filesforprogress.org/datasets/2021/1/dfp_vox_election_trust.pdf>). Additionally, forty-two percent of Independents said they did not trust the election results. Tim Mallow and Doug Schwartz, '74% of voters say democracy in the US is under threat, Quinnipiac University national poll finds; 52% say president trump should be removed from office' (*Quinnipiac University*, 11 January 2021) <https://poll.qu.edu/images/polling/us/us01112021_usmk38.pdf>.
[171] 'Majority Back Capitol Riot Commission' (*Monmouth University Polling Institute,* 17 March 2021) <https://www.monmouth.edu/polling-institute/reports/MonmouthPoll_US_031721/> accessed 8 July 2021.
[172] ibid.
[173] Chris Jackson and Kate Silverstein, 'Over half of Republicans believe Donald Trump is the actual President of the United States' (*IPSOS,* 21 May 2021) <https://www.ipsos.com/en-us/news_and_polls/over-half-republicans-believe-donald-trump-actual-president-united-states> accessed 8 July 2021. For this survey, a sample of 2,007 Americans age 18+ from the continental United States, Alaska, and Hawaii were interviewed online in English. The sample includes 909 Democrats, 754 Republicans, and 196 independents. Weighting was then

reflect the success of the disinformation campaign led by former President Trump regarding the integrity of the 2020 election.

While the insurrection was ultimately quashed by law enforcement, and Congress subsequently certified the Electoral College results,[174] the disinformation that inspired the insurrection remains at the forefront of American public discourse as former President Trump remains a pivotal figure in the Republican Party and Republican legislators at the federal and state levels are using these lies as a pretext for introducing an unprecedented number of bills directed to making it more difficult for Americans to vote. As of March 2021, legislators had introduced 389 bills with restrictive provisions in 49 states.[175] This coordinated attack on one of the most fundamental elements of democratic citizenship demonstrates the dangers of disinformation campaigns led by state actors, as well as the enduring viability of such campaigns.

2. Ignoring the (Red) Elephant in the Room

While the federal government is engaged in affirmative efforts to address disinformation, these efforts are primarily directed to disinformation from foreign actors, including Russia and China. While addressing the threats posed by foreign actors is an integral part of broader strategies to counter the dangers posed by disinformation in the digital age, these efforts ignore the significant role that domestic state actors play in disseminating disinformation and that such disinformation stems primarily from the Republican Party and its allies.

The American government's approach to combatting disinformation campaigns includes the establishment of special units whose focus is to counter foreign influence and share threat

employed to balance demographics to ensure that the sample's composition reflected that of the adult population according to Census data and to provide results intended to approximate the sample universe. The precision of Ipsos online polls is measured using a credibility interval. In this case, the poll is accurate to within ± 2.5 percentage points, 19 times out of 20 in the event that all Americans had been polled.

[174] It is important to note, however, that 145 Republican member of the House refused to certify the electoral results in particular swing states based on the Trump led disinformation campaign and at the time of writing, Republican at the federal and national level continue to challenge the results of the 2020 election.

[175] 'State Voting Bills Tracker: State lawmakers continue to introduce voting and elections bills at a furious pace' (*Brennan Center for Justice*, 28 May 2021) <https://www.brennancenter.org/our-work/research-reports/state-voting-bills-tracker-2021 > accessed 18 June 2021.

information with the private sector.[176] In 2016, the federal government established the Global Engagement Center, which is charged with leading the government's response to counter propaganda and disinformation from international terrorist organisations and foreign countries. In the fall of 2017, the FBI established the Foreign Influence Task Force to identify and counteract malign foreign influence operations targeting the United States.

A 2019 paper produced by the Combat Targeted Disinformation Campaigns team, operating under the auspices of the DHS's Analyst Exchange Program, proposes several actions for a variety of stakeholders to combat online disinformation. The paper defines a 'disinformation campaign' as occurring 'when a person, group of people, or entity (a "threat actor") coordinate to distribute false or misleading information while concealing the true objectives of the campaign'.[177] The target of such a campaign is the person or group the threat actor aims to influence in order to achieve the campaign's objective.[178] None of the Combat Targeted Disinformation Campaigns team's recommendations involve addressing the problem of domestic disinformation campaigns from state actors.[179] Additionally, the report provides no analysis or discussion of domestic state-sponsored disinformation campaigns and the dangers flowing therefrom.

Thus, while the United States continues to develop a broad and multi-layered strategy for countering foreign disinformation campaigns, there is a dearth of measures directed to addressing domestic disinformation from state actors and a lack of recognition of the overwhelmingly partisan nature of the current epistemic crisis in American public discourse. This sustained focus on foreign threats, notwithstanding the overwhelming evidence that domestic disinformation, particularly from state actors, poses a greater threat to American democracy ignores the political reality in the contemporary United States. This reality is that one of the two major political parties is attempting to wrest control of public discourse and the levers of power by disseminating disinformation to the masses and improperly manipulating the electoral process.

---

[176] 'Combatting Targeted Disinformation Campaigns: A whole-of-society issue' (*DHS*, October 2019) <https://www.dhs.gov/sites/default/files/publications/ia/ia_combatting-targeted-disinformation-campaigns.pdf> accessed 3 March 2021.
[177] ibid 4.
[178] ibid.
[179] The case studies of state-sponsored disinformation campaigns in the report are both foreign (Russia and China).

Conclusion

The doctrinal questions that inform this chapter concern how European and American free speech frameworks treat expression from state actors in public discourse, to what extent courts distinguish between private and official speech, and how such distinctions impact upon the treatment of, and protection afforded to, such expression. Using Arendt's reflections on the Pentagon papers as a springboard, this chapter interrogated the sufficiency of relevant doctrines in protecting the public's right to unmanipulated factual information in the digital age through an examination of domestic disinformation from state actors in Europe and the United States. The contemporary information ecosystem is markedly different from that which existed when Arendt provided her reflections on the Pentagon papers. While in prior centuries public discourse in democratic countries was facilitated by the free press and mass media, the primacy of traditional media - including national newspapers and local news stations - is under threat as more and more people turn to the internet, particularly social media platforms, for political news. At the same time, the meteoric increase in information sources has eroded the traditional media's historical gatekeeping function, including filtering information through professional editing, fact checking, and control through the political elite.[180] The internet has also fundamentally altered the way in which governments communicate with citizens as social media is a popular medium by which state actors routinely speak directly to the public.

Arendt observed that the one of the chief issues raised by the Pentagon papers was deception in the form of a 'deliberate lie'.[181] Deception is at the core of disinformation and may be described as its defining characteristic. Indeed, disinformation is so dangerous precisely because its primary purpose is to misinform the public for the purpose of eroding trust in democratic institutions and in traditional media sources. For this reason, disinformation may be regarded as distinct from the other forms of extreme speech examined in this thesis. For while 'hate speech' and terrorist-related expression create varying degrees of harm, they are clear examples of political speech, albeit in its ugliest forms. Disinformation, in contrast, particularly disinformation from state actors, is not imbued with basic attributes of political speech because its objective is not to share genuinely held beliefs - however unpopular or hateful - on matters of public interest and debate but, rather, to mislead others into believing something that the speaker knows to be objectively false in order to advance a political agenda.

[180] Fisher and Gaber (n 126) 5 - 6.
[181] Arendt (n 1) 4, 14.

For this reason, disinformation is antithetical to the primary principles that underpin European and American free speech frameworks. These include that free speech serves as an essential function of democracy as a means of holding state actors accountable to the people,[182] recognition of the vital role of the media in facilitating and fostering the public's right to receive accurate information regarding matters of public concern and debate,[183] and enabling 'informed decision-making by the electorate'.[184]

Additionally, disinformation from state actors is a particularly dangerous form of extreme speech because unlike private actors, even highly influential ones, state actors carry with them the imprimatur of the state in addition to the bully pulpit. Any argument that there is no meaningful difference between a state actor such as a president, prime minister or legislator using social media to speak directly to the public on matters of public concern, and a private citizen using social media for the same purpose, strains credulity. The relevant question is therefore whether in an era in which there is unprecedented direct communication between state actors and the public in the absence of the gatekeeping functions served by traditional media, whether free speech frameworks are effectively protecting the right of citizens to unmanipulated factual information.

How does examining recent developments in this area in Europe and the United States help to answer Arendt's question? First, this examination highlights the general lack of consideration in the American framework for the fact that a speaker's position or status may largely determine how a message is received and its potential impact on public discourse. This may be attributable, in part, to the American allegiance to the marketplace of ideas, which rejects the notion of an 'antidistortion interest' in preventing certain speech actors from gaining an unfair advantage in the political marketplace.[185] The relevant doctrines reflect a lack of concern from the USSC that an expansive approach to free speech may function to imperil the democratic norms and institutions the First Amendment is intended to strengthen and protect. Moreover,

[182] See *Citizens United v FEC*, 558 US 310, 339 (2010).
[183] See, e.g., *Magyar Helsinki Bizottsag v Hungary*, App no 18030/11 (ECtHR, 8 January 2016) para 165 (emphasising the 'vital role of the media in facilitating and fostering the public's right to receive and impart information and ideas has been repeatedly recognised by the Court'.).
[184] *Pickering v Bd of Educ of Township High Sch Dist 205*, 391 US 563 (1968).
[185] See, e.g., *Citizens United* (n 182) (in which the USSC rejected the underlying rationale for the antidistortion interest, which was to prevent corporations from obtaining 'an unfair advantage in the political marketplace' by using 'resources amassed in the economic marketplace' (883), reasoning said interest interfered with the 'open marketplace' of ideas (908) and that the government has no interest in 'equalizing the relative ability of individuals and groups to influence the outcome of elections.' (920)). The extent to which America's allegiance to the marketplace of ideas remains tenable in the digital age is explored in Chapter 6.

the federal government's efforts to address the dangers of disinformation ignore the most dangerous form of this type of extreme speech - domestic disinformation from state actors.

While the United States may be in a unique position due to the extent to which disinformation has become fully integrated into national politics, domestic disinformation from state actors poses a significant challenge to most contemporary democracies. The European framework may be better able to address this challenge because of its built-in safeguards against democratic backsliding. For example, it has recognised that even in the context of parliamentary immunity there are limits to freedom of expression, which serve as a restraint on government abuse and anti-democratic rhetoric. Additionally, it may be instructive to look to weakened European democracies for guidance. For example, in Hungary, disinformation arguably served to normalise and mainstream extreme-right beliefs and language, moving extreme-right ideology and the rhetoric that are acceptable in public discourse further to the right.

Based on the foregoing discussion, the answer to Arendt's question is that the interpretation of the First Amendment by the USSC has not only failed to protect the right to unmanipulated factual information, but that the American approach to free speech has arguably functioned to subvert this right. While Europe is sounding alarm bells regarding the problem of domestic disinformation from state actors, the United States is largely framing the problem as purely one to do with foreign adversaries, and American politicians are largely treating the current attacks on democracy from the Republican Party and manipulation of social media to effectuate anti-democratic ends as business as usual.

Liberal democracy is predicated on the principles of the public's right to access to accurate information and electoral accountability as democratic safeguards against tyranny and corruption.[186] Moreover, strong democracies require healthy information systems where citizens are able 'to come together to debate, discuss, deliberate, empathize, make concessions and work towards consensus'.[187] The primary function of free speech is to strengthen democracy by facilitating an informed populace and fostering a strong press that holds the government to account. This chapter considers the ways in which state actors use free speech

[186] Fisher and Gaber (n 126) 2.
[187] Benkler 'Mail-In Voter Fraud' (n 161) 2.

as a powerful weapon to misinform the public for anti-democratic ends. Accordingly, it is time to start thinking about this problem in terms of the positive right of the public to receive accurate and factual information in public discourse, and the ways in which governments, in particular the United States, are not only failing to protect this right but are actively undermining it and, in so doing, weakening the democratic order. Chapter 6 proffers that it is possible to locate a positive right of the public to access unmanipulated factual in American free speech jurisprudence.

Additionally, in examining the ways in which governments and political parties have leveraged social media to pollute the digital information ecosystem, the Oxford Disinformation Report emphasises that while many of the issues underlying the spread of disinformation, including polarisation, distrust, and the weakening of democratic norms and institutions predate the internet, and that the manipulation of social media is of concern, so too are many of the long-standing challenges facing democracies around the world.[188] This highlights important themes in this thesis, including the importance of taking a contextualised view of the challenges posed by extreme speech in the digital age and to appreciate the ways in which technological advances often exacerbate - rather than create - deeply rooted societal problems and tensions. Governments are quick to blame social media and defend increasingly aggressive efforts to control how platforms moderate content while avoiding the more challenging work of addressing the underlying problems that advances in technology exacerbate. This sustained focus on digital intermediaries allows governments to use the dangers of extreme speech as a justification for more onerous regulations that impact freedom of expression while step-stepping their own complicity in such problems.

Finally, as with the other types of extreme speech considered in Part II, the problem of domestic disinformation from state actors must form part of a much larger and more nuanced conversation regarding the appropriate limits on free speech in the digital age. In order for this conversation to be productive, it must include all relevant actors in the contemporary information ecosystem, including the traditional media, social media platforms, governments and supranational bodies. Chapter 6 provides a proposal for the proper scope of this

[188] Bailey (n 8) 21.

conversation, along with other reflections and recommendations that emerge from the examinations undertaken in the preceding chapters.

Chapter 6

Reflections and Recommendations From a Comparative Analysis of the European and American Approaches to the Regulation of Extreme Speech[1]

'*While we now may be coming to the realization that the Cyber Age is a revolution of historic proportions, we cannot appreciate yet its full dimensions and vast potential to alter how we think, express ourselves, and define who we want to be. The forces and directions of the Internet are so new, so protean, and so far reaching that courts must be conscious that what they say today might be obsolete tomorrow'*.[2]

Introduction

The preceding chapters in Part II, supported by the doctrinal and theoretical foundation provided in Part I, examine the challenges and dangers posed by 'hate speech', terrorist-related expression, and disinformation in the digital age through a comparative analysis of the European and American approaches to the regulation of extreme speech. This, the final chapter, offers some reflections and recommendations based on these examinations, which enrich our understanding of the potential benefits and pitfalls of regulatory efforts in these areas while illuminating the harms flowing from these types of extreme speech. In so doing, this chapter aids the primary objective of this thesis, which is to offer insights into the broader discourse concerning the appropriate limits on freedom of expression in liberal democracies in the digital age.

This chapter proceeds as follows: Part I relies on the philosophical work undertaken in Chapter 2 in proposing that the philosophical underpinnings of the American approach warrant re-

[1] Chapter 6 is derived in part from an article written by the author published in *Communications Law*: Eliza Bechtold, 'Has the United States' Response to the COVID-19 Pandemic Exposed the Marketplace of Ideas as a Failed Experiment?' (2020) 25(3) *Communications Law*, 150 -160.
[2] USSC Justice Anthony Kennedy, writing for the majority in *Packingham v North Carolina*, 137 SCt 1730, 1736 (2017).

examining in the digital age and following the presidency of Donald Trump. It argues that the disinformation campaign concerning the 2020 presidential election (examined in Chapter 5) strongly suggests that the primary principles upon which theories of democratic legitimacy rest are no longer tenable. Part I further contends (drawing from the examinations undertaken in Chapters 1, 3, and 5) that the digital age has exposed the marketplace of ideas, the cornerstone of America's free speech paradigm, as a failed experiment. Part II considers the question of whether certain free speech doctrines, largely developed in the twentieth century, remain fit for purpose in the digital age. These include the USSC's government speech doctrine as well as European and American doctrines relating to immunity for state actors (all of which are considered in Chapter 5). Part III engages with human rights penality in Europe (drawing primarily from Chapter 3) and reflects on the broader lessons from recent efforts to criminalise extreme speech. Part IV (relying on the examinations of the ECtHR's approach in Article 10 cases undertaken in Chapters 1, 3, and 4) argues that the ECtHR's methodology in the context of extreme speech would benefit from greater conceptual clarity and coherence. Part V (relying primarily on the analysis of recent efforts to regulate online terrorist-related expression in Europe provided in Chapter 4) reflects on the broader dangers to freedom of expression resulting from aggressive efforts to regulate digital intermediaries. Part V also considers the extent which efforts at the state and national levels in the United States portend an increasingly entangled relationship between the government and digital intermediaries, and the potential implications for the state action doctrine (discussed in Chapter 1) flowing therefrom.

I. The Philosophical Underpinnings of the American Approach to Extreme Speech Warrant Re-examining in the Digital Age

Chapter 2 examines the primary philosophical considerations that inform this thesis and considers, among other things, the link between theories of democratic legitimacy and the expansive protections afforded to extreme speech in the American constitutional framework. Section A below relies on this philosophical groundwork and the analysis of the disinformation campaign targeting the 2020 American presidential election in Chapter 5 to argue that it is time to re-examine America's predominant reliance on theories of democratic legitimacy. Section B posits that recent events in the United States, particularly those highlighted in Chapters 3 and 5, suggest that the marketplace of ideas is a failed experiment.

A. Reliance on Theories of Democratic Legitimacy to Justify Expansive Protection of Extreme Speech in American Public Discourse May No Longer Be Tenable

As discussed in Chapter 2, theories of democratic legitimacy are grounded in notions regarding the appropriate relationship between the individual and the state, and are premised on the principle that allowing citizens the opportunity to meaningfully participate in public discourse is a necessary precondition for a legitimate democracy.[3] For example, Dworkin posits that it is illegitimate for the government to impose limitations on expression because it deprives citizens of a fair opportunity to express their attitudes and opinions in public discourse. He reasons that '[i]f we expect bigots to accept the verdict of the majority once the majority has spoken, then we must permit them to express their bigotry in the process whose verdict we ask them to respect'.[4] Heinze argues that restrictions on 'hate speech' in a longstanding, stable, and prosperous democracy (LSPD) such as the United States are unwarranted because 'as a matter of social and civic awareness, plurality of opinion is, in LSPDs, robust enough to enable counter-speech and scrutiny of hate speakers and groups'.[5] It is for this reason, Heinze contends, 'that prohibitionists have failed to substantiate the kinds of empirical links from speech to violence or to discrimination that can more easily be adduced within non-LSPDs'.[6] While Heinze's theory is limited to 'hate speech' it is relevant to the broader point made here, which is that theories linking the legitimacy of a democracy to the expansive protection of extreme speech warrant re-examining in the digital, post-Trump era in the United States.[7]

In commenting on the relationship between democratic legitimacy and free speech in 2017, Frederick Schauer observed that while he did not believe that 'claims about the relationship between an opportunity to object and the incidence of compliance [with the law] have been falsified', at the same time, such claims 'have not been established, and the research that might be understood to establish them turns out not to do so'.[8] Schauer further notes that it is important 'to distinguish a description of belief from a description of behaviour' and that '[e]ven if it is the case that freedom of speech increases the degree of belief in the legitimacy

[3] See Chapter 2, pp 61 - 64.
[4] ibid 4 (Ronald Dworkin, "Even Bigots and Holocaust Deniers Must Have Their Say," *The Guardian* (14 February 2006) <http://www.theguardian.com/world/2006/feb/14/muhammadcartoons.comment> accessed 10 March 2019).
[5] ibid 6 (Eric Heinze, *Hate Speech and Democratic Citizenship* (OUP 2016)).
[6] ibid.
[7] ibid.
[8] Frederick Schauer, 'Free Speech and Obedience to Law' (2017) 32 Const Comment 661, 673.

of law, and belief in the obligation to obey disagreeable laws, on the question of whether freedom of speech increases actual compliance with laws with which people disagree, the best we can do is simply to say that we do not know'.[9] This section argues that the fall-out from the 2020 presidential election, in particular, the Capitol insurrection, provide persuasive evidence that the answer to Schauer's question is most likely no. Thus, it is an opportune time to re-examine the relationship between freedom of expression and democratic legitimacy in the American constitutional framework, as philosophical arguments should be tested by real world examples from time to time in order to establish their continued relevance and viability.

As highlighted throughout this thesis, the United States provides the most expansive protections to extreme speech in public discourse of any contemporary liberal democracy. Thus, it is an ideal framework to test democratic legitimacy as a philosophical justification for broad free speech protections in the digital age. In the contemporary United States, bigots are free to express their views on matters of public concern with very limited exceptions and may freely express opinions and ideas that are demeaning on the basis of race, ethnicity, gender, religion, and other similar grounds.[10] They may also advocate for the use of terrorism and other types of violence so long as such expression does not satisfy the strict requirements of the *Brandenburg* test.[11]

The peaceful transfer of power from one head of government to the next following a free and fair election is one of the most essential components of a functioning democracy. As discussed in Chapter 5, the 2020 election was one of the most secure elections in American history, with no evidence of significant voter fraud or other irregularities.[12] Freedom House's 2020 report on the United States gave the American electoral process 10 out of 12 points.[13] Thus, the 2020 election offers the ideal test case for theories of democratic legitimacy in the digital age. This is a two-part inquiry: first, what does the available evidence suggest regarding whether those who disagree with the outcome of the 2020 election accept the result? And second, do those

[9] ibid.
[10] See the discussion of the American approach to 'hate speech' in Chapter 3, pp 97 - 100 and the discussion of exceptions to free speech protections in the United States in Chapter 1, pp 35 - 40.
[11] See Chapter 4, p 151.
[12] See Chapter 5, p 183.
[13] The electoral process scale is one of seven that comprise Freedom House's overall ratings of countries as 'free', 'partly free', or 'not free'. It includes three areas: whether the head of the government is selected through free and fair elections; whether national legislators are chosen through free and fair elections; and whether a country's electoral laws and framework are fair. See 'Freedom in the World 2020: United States (*Freedom House*, 2020) <https://freedomhouse.org/country/united-states/freedom-world/2020> accessed 2 July 2021.

who refuse to accept the result nevertheless feel an obligation to respect it? This section addresses each of these questions in turn.

Chapter 5 highlights the extent to which the online disinformation campaign led by former President Trump regarding the integrity of the 2020 election was a resounding success. Polling conducted just after the election and for several months following President Biden's inauguration revealed similar attitudes among the American public, with approximately one third believing that President Biden's victory was the result of widespread voter fraud and the majority of Republicans refusing to accept that the legitimacy of Biden's presidency on this basis. [14] Given the context in which the 2020 election occurred - in a country that provides expansive protections to speech in public discourse, followed by the most secure election in American history, and one of the greatest margins of victory in any election in the history of the country - the percentage of Americans that do not accept the results of the election is staggering. Dworkin argues that we must be willing to tolerate extreme speech so that those who disagree with laws 'accept the verdict of the majority once the majority has spoken'.[15] It is therefore reasonable to inquire as to whether society must continue to tolerate extreme speech when it becomes apparent, as is the case in the contemporary United States, that so many of those who disagree with the majority's verdict refuse to accept it.

The second part of this inquiry examines whether those who refuse to accept the legitimacy of President Biden's victory feel an obligation respect it. Clearly, those who stormed the Capitol in an effort to overthrow American democracy based on the false belief that the election was stolen from former President Trump refused to respect such legitimacy. While significant, the more relevant question is the proportion of the American public that believes in the legitimacy of those efforts. In polling conducted in April and May of 2021, twenty-one percent of Republicans reported believing that the attack on the Capitol was justified[16] and twenty-eight percent of Trump voters expressed a favourable opinion of the insurrectionists,[17] with thirty-

[14] See Chapter 5, pp 184 - 186.
[15] Dworkin (n 4).
[16] Fourteen percent of total respondents and only seven percent of Democrats believed the attack was justified (YouGov, 'Yahoo! News COVID-19 Vaccination Survey - 20210526 (24-26 May 2021)' <https://docs.cdn.yougov.com/zjdg6ujrzh/20210526_yahoo_vaccine_tabs.pdf> accessed 2 July 2021.
[17] YouGov, '*The Economist*/YouGov Poll, 22-25 May 2021' <https://docs.cdn.yougov.com/pxuc7wjg52/econTabReport.pdf> accessed 2 July 2021.

three percent describing these individuals as 'patriots'.[18] Additionally, more than thirty percent of Republicans rejected the basic premise of democracy that 'the loser in an election must concede defeat'.[19] A February 2021 survey by the American Enterprise Institute found that nearly three in ten Americans completely or somewhat agreed with the statement that 'if elected leaders will not protect Americans, the people must do it for themselves even if it requires taking *violent* actions'.[20]

Thus, post-election polling suggests that there are a significant number of Americans, particularly Republicans, who not only refuse to accept the legitimacy of Joe Biden's presidency but who also believe that political violence, both during the Capitol insurrection and more generally, is justified in the contemporary United States. Beliefs regarding the integrity of the 2020 election and the use of political violence to address the perceived injustices resulting therefrom may and should be linked to the disinformation campaign that created and fomented false information regarding the legitimacy of the American electoral system.[21] This provides compelling evidence that an expansive approach to free speech does not increase compliance with laws with which people disagree. On the contrary, it suggests that extreme speech, particularly in the form of online disinformation from state actors, may *increase* unlawful behaviour as well as foment democratic backsliding.

The events following the 2020 election also suggest that it is reasonable to question Heinze's assumption regarding restrictions on 'hate speech', that is, that such restrictions are unwarranted because LSPDs are stable enough to enable counter-speech and scrutiny of those who disseminate such expression. The fall-out from the Capitol insurrection and related

[18] University of Massachusetts, 'University of Massachusetts Amherst/WCVB National Poll of President Biden's First 100 Days' <https://polsci.umass.edu/sites/default/files/Biden100DaysToplines%20%281%29.pdf> accessed 2 July 2021. In this poll, seventeen percent of Independents and five percent of Democrats also selected this description.

[19] Only eight percent of Democrats share this view. See Ipsos, 'Ipsos/Reuters Poll: The Big Lie – Over half of Republicans believe Donald Trump is the actual President of the United States' <https://www.ipsos.com/sites/default/files/ct/news/documents/2021-05/Ipsos%20Reuters%20Topline%20Write%20up-%20The%20Big%20Lie%20-%2017%20May%20thru%2019%20May%202021.pdf> accessed 2 July 2021.

[20] Daniel A Cox, 'After the ballots are counted: Conspiracies, political violence, and American exceptionalism', (*Survey Center on American Life,* 11 February 2021) <https://www.americansurveycenter.org/research/after-the-ballots-are-counted-conspiracies-political-violence-and-american-exceptionalism/> accessed 2 July 2021 (emphasis added).

[21] By way of example, during the Congressional investigation into the Capitol insurrection, Capitol police officers testified that they heard individuals attacking the Capitol claiming, among other things, that they were there to 'stop the steal' and that 'President Trump sent us'. See Charlotte Alter, 'The Start of the Jan. 6 Insurrection Inquiry Shows its Stakes – And its Shortcomings' (*Time.com,* 27 July 2021) <https://time.com/6084466/jan-6-inquiry-begins-capitol-insurrection> accessed 27 July 2021.

polling provide empirical links between extreme speech and a level of violence directed to the government that is rarely observed in stable democracies. This suggests that it is time to revisit longstanding assumptions regarding the relationship between the regulation of extreme speech and the strength and stability of established democracies.

B. The Contemporary Information Ecosystem has Exposed the Marketplace of Ideas as a Failed Experiment

As discussed in Chapter 3, the marketplace of ideas, which emerged during the early twentieth century, serves as the cornerstone of American's contemporary free speech paradigm.[22] It is premised on the principle that public discourse must remain free and open in order to prevent the government from controlling the search for 'political truth'.[23] The USSC warns that the absence of such a marketplace will result in a public that is unable to choose a government that reflects its informed will.[24] A free and open marketplace of ideas serves as one of the primary justifications for the expansive protections afforded to extreme speech in the American constitutional framework.

The metaphor emerged during a period when limited information pathways existed, both with respect to the ways in which individuals communicated with one another and how the government communicated with the public. Its underlying presumption - that there is a single 'marketplace' in which all voices operate - is untenable in the contemporary information ecosystem. Chapter 5 highlights the extent to which the public discourse in the United States is fragmented and polarised, with Democrats and Republicans existing in 'two nearly inverse news media environments'.[25] The disparities in the beliefs of Democrats and Republicans regarding the integrity of the 2020 election provide disturbing confirmation of this fact.

Additionally, the marketplace of ideas does not account for the fact that the government itself plays an important role as a speech actor in public discourse, and the ways in which the digital age has transformed the way that state actors communicate with the public. As highlighted in Chapter 5, in 2020 alone, nearly all American legislators posted on official social media

[22] See Chapter 3, pp 112 - 115.
[23] ibid 22 (*Consolidated Edison Co v Public Serv Comm'n*, 547 US 530, 538 (2011)).
[24] ibid (*Sorrel v IMS Health, Inc*, 564 US 552, 583 (2011)).
[25] See Chapter 5, p 163 (Mark Jurkowitz and others, 'U.S. Media Polarization and the 2020 Election: A Nation Divided' (*Pew Research Center,* 24 January 2020) <https://www.journalism.org/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/> accessed 8 June 2020).

accounts and all of these state actors used Facebook.[26] The clearest example of this phenomenon is former President Trump, who used Twitter to communicate to the masses on a daily basis. His allies, often through social media, continue to pollute public discourse with disinformation that is used as a pretext for onerous voting restrictions that further weaken American democracy.[27]

That so much expression in contemporary public discourse takes place on social media merits consideration of the ways in which content is amplified and targeted via and on these platforms. In particular, the way in which user behaviour in conjunction with the ads-based business models of digital intermediaries results in platforms amplifying content that keeps users engaged, including content that is harmful, misleading, and extreme.[28] This means that certain voices are amplified while others are stifled. While the content moderation decisions of private actors such as digital intermediaries do not implicate the First Amendment,[29] these decisions are increasingly relevant to how the public accesses information, including information regarding official acts of government. The point here is that in the digital age technology largely dictates what speech is amplified on the forums in which so much of the public receives information regarding matters of public concern.

The examinations undertaken in Chapters 3 through 5 demonstrate that the ultimate objective of the marketplace of ideas - to reach truth through the competition of ideas - is premised on outdated assumptions, including that there is a single marketplace to which all voices have equal access, that all citizens agree that objective truth exists, and that there is an easily discernible difference between facts and opinions. Additionally, a fundamental part of America's free speech tradition is the principle that an informed public is the essence of working democracy.[30] Yet, the USSC seemingly assumes that the marketplace of ideas will function effectively so long as the government is not proscribing certain private expression based on content or viewpoint, which will then produce an informed electorate who relies on

[26] 'The world leaders with the most Twitter followers 2020' (*Statista*, 27 January 2021) <https://www.statista.com/statistics/281375/heads-of-state-with-the-most-twitter-followers/> accessed 2 May 2021.
[27] See relevant discussion in Chapter 5, pp 199 - 200.
[28] Daphne Keller, 'Amplification and Its Discontents', (*Knight First Amendment Institute*, 8 June 2021) <https://knightcolumbia.org/content/amplification-and-its-discontents> accessed 6 July 2021.
[29] The potential First Amendment implications of excessive government regulation of intermediaries is examined infra in Part V.
[30] See *Minneapolis Star v Minnesota Comm'r*, 460 US 575, 585 (1983) (quoting *Grosjean v. Am Press Co*, 297 US 233 (1936)).

'truth' in casting ballots, which will ultimately lead to a government that effectively represents the will of the people. These assumptions do not withstand scrutiny in American public discourse in the digital age, as state actors and political figures are actively endeavouring to misinform the public on vital issues of public concern in an environment in which disinformation is communicated to the masses with a single keystroke. This disinformation extends beyond the integrity of the 2020 election to other important issues, such as the COVID-19 pandemic and efforts to subvert the important First Amendment role of the press in serving as a check on government power. This reveals a contemporary tension between the American approach to free speech and the health of American democracy.

At present, it is clear that there is no single contemporary online 'marketplace', and that American public discourse is manipulated by disinformation and other forms of extreme speech, with the amplification of certain voices over others. What does all of this mean for the marketplace of ideas? At the very least, it means that this is a time for serious introspection concerning whether it is time to abandon the cornerstone of America's free speech paradigm. To do so would be a significant undertaking, as the marketplace of ideas is interconnected with other constitutional principles, including the government speech doctrine, the conceptualisation of liberty in negative terms, and the presumption that viewpoint and content-based proscriptions on speech are presumptively unconstitutional. Notwithstanding these challenges, it is important to consider the ways in which the current approach has not only permitted, but arguably fostered, an environment in which state actors utilise social media and other online platforms to effectuate anti-democratic ends.

II. Are Twentieth Century Free Speech Doctrines Fit For Purpose in the Digital Age?

Another question that emerges from this CCL inquiry is the extent to which advancements in technology render elements of existing free speech doctrines, largely developed during the prior century, unfit for purpose. In his concurrence in a recent decision involving a case in which former President Trump blocked users from his official Twitter account, USSC Justice Clarence Thomas alluded to this question by highlighting '[t]he principal legal difficulty that surrounds digital platforms - namely, that applying old doctrines to new digital platforms is rarely straightforward' and predicting that the USSC 'will soon have no choice but to address

how our legal doctrines apply to highly concentrated, privately owned information infrastructure such as digital platforms'.[31]

Justice Thomas' comments relate to Frederick Schauer's observation that 'legal doctrine is created in light of empirical estimates about the state of the world, and as new events cause us to revise our previous estimates, it should come as little surprise that these new events should cause us to re-examine the doctrines that have emanated out of earlier and possibly outdated empirical estimates'.[32] The ways in which advances in technology, particularly the advent of social media, have transformed how private and state actors participate in public discourse should cause us to re-examine the extent to which certain doctrines remain fit for purpose, that is, whether such doctrines reflect outdated empirical estimates. This section offers some recommendations regarding specific doctrines that warrant such re-examination in the digital age, which build on the doctrinal foundation laid in Chapter 1 and the doctrinal examinations undertaken in Chapters 3 and 5 concerning 'hate speech' and disinformation, respectively.

A. The Scope of Immunity for State Actors Should Not Extend Beyond Traditional Conceptions

Chapter 5 tackles the dangers posed by online disinformation from state actors and highlights that traditional conceptions of immunity for the expression of these actors derive from long-standing principles, including the value of free and open debate among lawmakers to democratic deliberation and that frank communication between state actors and the public strengthen democracy. These principles developed in the pre-digital age, during which fewer information sources existed and the traditional media could more effectively exercise its gatekeeping function. In contrast, in the contemporary information ecosystem state actors communicate directly to the public on a routine basis, via social media and other forums for online speech, with traditional media playing a secondary role. This makes it impracticable for the traditional media to filter information through professional editing and fact-checking. This shift represents a marked departure from the ways in which state actors communicated with the public in the pre-digital era and the traditional ways in which the media fulfilled its

[31] *Biden v Knight First Amendment Inst*, 141 SCt 1220, 1221 (2021). The Supreme Court vacated the decision of the Second Circuit Court of Appeals due to the change in presidential administration.
[32] Frederick Schauer, 'Is it Better to Be Safe than Sorry? Free Speech and the Precautionary Principle' (2009) 36(2) Pepperdine L Rev 301, 307.

important free speech role of facilitating the public's right to receive accurate information regarding matters of public concern and holding state actors to account.

In the United States, these changes have led to state actors advocating for increasingly broad interpretations of immunity within statutory and doctrinal frameworks. Federal courts have proven amenable to such claims, holding that posts on social media fit within the broad range of duties elected officials perform for constituents and, as a result, fall within the American statutory immunity framework. The DOJ has gone so far as to argue that the Westfall Act may cover *criminal* conduct that is prompted, in part, by *personal* motives.[33] Given the extent to which contemporary American discourse is polluted by extreme speech from state actors, arguments for broadening immunity for state actors merit demanding scrutiny that, at present, American courts appear unwilling to apply.

While the ECtHR's jurisprudence in this area is lacking in conceptual clarity, it evidences a recognition by the Court of the inherent challenges associated with delineating the proper scope of protection for the expression of state actors in public discourse, and the tension between the notion that state actors have heightened duties and responsibilities as a result of their outsized influence in public discourse and the longstanding principle that freedom of speech is especially important for state actors (in particular, legislators).[34] Additionally, the Committee on Human Rights report highlighted in Chapter 5 raises valid concerns regarding the importance of a wide interpretation of freedom of speech for parliamentarians who belong to opposition parties and whose ideas differ from those in the majority, while acknowledging that 'hate speech' and calls for violence from those protected by parliamentary immunity should not be tolerated. This demonstrates that while American courts are largely ignoring the dangers and challenges associated with extreme speech from state actors, the ECtHR and Council of Europe are attempting to grapple with the difficult regulatory questions that arise in this area.

A broader intra-jurisdictional discourse would therefore be particularly instructive in this area, as the basic protections provided within immunity frameworks do not differ markedly in liberal democracies. For example, the principle of heightened duties for state actors as speech actors, as opposed to 'regular' citizens, is not foreign to the United States Constitution. The USSC's

[33] See Chapter 5, pp 166 - 167.
[34] See discussion of the ECtHR's approach to parliamentary immunity in Chapter 5, pp 166 - 170.

decision in *Wood v Georgia* is relevant for its suggestion that on a different set of facts the Court could resolve, consistent with precedent, that restrictions on expression from state actors do not contravene the First Amendment. Specifically, the Court suggested that speech that interferes with the performance of a state actor's 'official' duties and/or creates a substantial likelihood that expression may 'disrupt the administration of justice' may be subject to legitimate restrictions.[35] This express recognition of the relevance of the status of state actors when speaking on matters of public concern could lend support to an argument that restrictions on anti-democratic expression from these actors, including disinformation attacking the integrity of democratic norms, does not contravene free speech protections. The relevant question is whether American courts, like their European counterparts, are willing to engage in such debates.

As highlighted above, the exponential increase in communication between the government and the public in the digital age has profoundly impacted the traditional media's ability to serve as an effective gatekeeper. This, coupled with the fact that one of the most significant threats to contemporary democracies is domestic disinformation from state actors, makes for a powerful argument that expanding traditional conceptions of immunity in the digital age undermines the original rationale for such immunity, which is to facilitate the operation of democracy.[36] While the changes to public discourse brought by the digital age warrant critical reflection regarding the extent to which existing free speech doctrines remain fit for purpose, such reflection in the context of parliamentary immunity suggests that efforts to expand the scope of immunity to non-traditional forms of communication undermine both the original rationale for such immunity and democracy more broadly by endowing state actors with expansive powers to spread disinformation and other forms of extreme speech with limited accountability.

B. The USCC's Government Speech Doctrine is a Twentieth Century Anachronism that Undermines Free Speech and Ignores the Important Role that State Actors Play in Public Discourse

The government speech doctrine is a uniquely American conception with no comparable principle in the European constitutional framework. This may be explained, in part, by Europe's permissive approach to content and viewpoint-based proscriptions on expression, as

[35] See discussion of *Wood v Georgia* in Chapter 5, pp 174 - 175.
[36] See ibid, p 165.

the premise of the government speech doctrine is that the government must take particular viewpoints and reject others when it engages in the process of governing.[37] While the premise of the government speech doctrine is sound in principle, it only makes sense if the government is speaking in ways that relate to its performance of traditional government functions, such as the promotion of public health measures or taxation, and in so doing, government officials conform to traditional democratic norms. Rising instances of 'hate speech' and disinformation from state actors in the United States, including racist and xenophobic attacks on immigrants and the campaigns targeting the integrity of the American electoral largely conducted via social media, suggest that it is time to re-examine whether the government speech doctrine remains fit for purpose.

While it is true that in order to govern, the government must 'say something', the USSC's jurisprudence reflects a presumption by the Court that whatever the government says will concern the business of governing and that such speech will not be wielded in a manner that improperly manipulates public opinion. It also presupposes that the portion of the public that disagrees with the government's message will express such disagreement by way of the ballot box. This relates to the inherent presumption underlying the doctrine that government speech will not serve to undermine American democracy because an informed populace and functional electoral process will serve as an effective safety valve against democratic backsliding. The disinformation campaign concerning the integrity of the 2020 election discussed in Chapter 5 exposed flaws in the basic presumptions upon which the government speech doctrine rests. This type of extreme speech from state actors is directed to undermining the public's faith in democratic processes and weakening democratic institutions, rather than to the business of governing. This is a distinction with an important difference that does not appear to be addressed, or even contemplated, in USSC's government speech doctrine jurisprudence.

Anti-democratic speech from state actors suggests that principles and assumptions underpinning government speech doctrine are no longer tenable. At least one federal court has questioned the implicit suggestion in government speech cases that the government is 'simply one more participant in the marketplace of ideas', arguing that this ignores 'the force of government, as compared to private speech and, even more importantly, the access that

[37] See Chapter 5, p 175.

government speech has to free media'.[38] And while the Supreme Court provides little guidance in terms of fleshing out the contours of the government speech doctrine, in particular, its relationship to the marketplace of ideas, other free speech principles are instructive. These include that ideas should compete in the marketplace without government interference,[39] that 'drowning out private sources of speech' in the marketplace constitutes improper government interference,[40] and that the government may not speak so loudly as to 'make it impossible for other speakers to be heard by their audience'.[41] These principles support the conclusion that the government speech doctrine is not fit for purpose in the digital age. While these cases address traditional forms of government speech, such as granting broadcasting licenses and state mandated advertisements, the espoused principles apply with equal force to online disinformation from state actors, through which they actively endeavour to control and manipulate the marketplace of ideas by, among other things, using social media to pollute public discourse and improperly manipulate public opinion.

If an informed populace is the most potent of all restraints upon misgovernment, surely a misinformed one paves the way for the absence of such restraints. The world has watched this play out in real time as the domestic disinformation campaign regarding the 2020 election undermined democratic institutions and inspired the most serious attack on American democracy in recent memory. Re-examining the government speech doctrine in the digital age reveals that the empirical estimates upon which it is based, including its implicit assumptions regarding the speech of state actors, are outdated. Whether and what to replace this outdated doctrine with are topics for future research.

C. The Fairness Doctrine: Locating a Positive Right to Access Unmanipulated Factual Information in American Free Speech Jurisprudence

As discussed in Chapter 1, while the European framework recognises positive obligations on behalf of governments to protect fundamental rights, the American framework embraces traditional conceptions of negative liberty and the principle that the First Amendment only

[38] *RJ Reynolds Tobacco Co v Bonta*, 272 FSupp2d 1085, fn 20 (ED Cal 2003). This argument is particularly compelling in the digital age.
[39] *Citizens United v FEC*, 130 SC 876, 906 (2009) (internal citations omitted).
[40] *NAACP v Hunt*, 891 F2d 1555, 1566 (11th Cir 1990).
[41] *Warner Cable Communications, Inc v City of Niceville*, 911 F2d 634, 638 (11th Cir 1990).

protects individuals from the government.[42] Thus, in the context of facilitating a healthy information ecosystem, European state actors are endowed with greater latitude in regulating in this space. However, a doctrine from the twentieth century may prove useful in contemplating ways to tackle the contemporary disinformation crisis threatening American democracy without undermining America's free speech paradigm.

The fairness doctrine was a Federal Communications Commission[43] (FCC) policy in effect from 1949 to 1987 that required television and radio stations to broadcast contrasting views in any discussion of controversial issues of public importance and to provide a right of reply to individuals who were attacked during such discussions.[44] The basic principle underlying the fairness doctrine was the First Amendment right of the public to be informed, rather than the right on the part of the government, any broadcast licensee, or any individual member of the public to broadcast his own particular views on any matter.[45] The USSC's jurisprudence concerning the constitutionality of the fairness doctrine is significant because it identifies a *positive* right of the public to access information.[46]

The USSC's fairness doctrine jurisprudence provides a normative foundation for a nuanced discussion regarding the regulation of the American media landscape grounded in a positive right of the public to unmanipulated factual information in public discourse. In 1969, the USSC upheld the constitutionality of the fairness doctrine in *Red Lion Broad. Co. v. FCC*. In this case, the Court recognised a First Amendment 'right of the public to receive suitable access to social, political, aesthetic, moral, and other ideas and experiences' through broadcast media.[47] The Court noted that before regulation of broadcast frequencies, a scarce resource at the time, 'the result was chaos'.[48] In view of the scarcity of broadcast frequencies and the government's role in allocating those frequencies, the Court concluded that '[w]ithout government control, the medium would be of little use because of the cacophony of competing

[42] See Chapter 1, p 46.
[43] The FCC regulates interstate and international communications by radio, television, wire, satellite, and cable in all fifty states, the District of Columbia and United States territories. An independent United States government agency overseen by Congress, the FCC is the federal agency responsible for implementing and enforcing America's communications law and regulations.
[44] Communications Act of 1934 (47 USCA) s 315(a).
[45] See *Columbia Broad Sys, Inc v Democratic Nat'l Comm*, 412 US 94, 112 (1973)
[46] See, e.g., ibid 101 (in which the USSC opined that '[b]alancing the various First Amendment interests involved in the broadcast media and determining what best serves the public's right to be informed is a task of a great delicacy and difficulty.').
[47] See *Red Lion Broadcasting Co v FCC*, 395 US 367, 390 (1969).
[48] ibid 376.

voices, none of which could be clearly and predictably heard'.[49] In the USSC's view, the 'thrust of restrictions like those in the fairness doctrine [on the broadcasting industry] has generally been to secure the public's First Amendment interest in receiving a balanced presentation of views on diverse matters of public concern'[50] and to promote the 'First Amendment goal of producing an informed public capable of conducting its own affairs'.[51] This is linked to other key First Amendment principles, including that '[p]ublic debate must be both unfettered and *informed*.'[52]

When the FCC withdrew the fairness doctrine in 1987, it took the position that it was unconstitutional, and the effective annulment of *Red Lion* has been assumed by a variety of government officials and scholars.[53] However, the USSC has not overruled *Red Lion* and, as noted by Tim Wu - Professor of Law at Columbia Law School and special assistant to President Biden for technology and competition policy at the National Economic Council - 'in the law, no doctrine is truly dead' and the significant changes to the information ecosystem in the digital age strengthen the constitutional case for laws explicitly intended to improve political discourse.[54]

Moreover, as emphasised by Victor Pickard, Professor of Media Policy and Political Economy at the Annenberg School for Communication, while the FFC's abandonment of the fairness doctrine was ostensibly an attempt to remove 'unnecessary regulations,' its action supports an argument that the government has no legitimate role in regulating media markets and protecting positive freedoms associated with accessing information in public discourse.[55] However, Pickard correctly observes that the USSC jurisprudence provides normative grounding for a focus on the public's *positive* right to hear a broad range of information versus the negative right to be free from governmental interference in accessing information.[56] He recommends that American public discourse should actively address the kind of media ecosystem required

[49] ibid.
[50] *FCC v League of Women Voters of California,* 468 US 364, 383 (1984). See also *Red Lion* (n 47) 368: [i]t is right of viewers and listeners, not the right of broadcasters, which is paramount.'
[51] ibid (*Red Lion*) 392.
[52] *First Nat'l Bank of Boston v Belotti,* 435 US 765, 782 (1978) (in which the Court cited Justice Powell's dissent in *Saxbe v Washington Post Co*, 417 US 843, 862-63 (1974)) (emphasis added).
[53] Tim Wu, 'Is the First Amendment Obsolete?' (2018) 117(3) Mich L Rev 547.
[54] ibid 577.
[55] Victor Pickard, 'The Strange Life and Death of the Fairness Doctrine: Tracing the Decline of Positive Freedoms in American Policy Discourse' (2018) 12 Int J Com 3434, 3447.
[56] ibid 3439.

in a democracy, noting that in the three decades following the FCC's rejection of the fairness doctrine, it has mandated increasingly weakened obligations on broadcasters in order to maintain their significant monopoly privileges.[57] In Pickard's view, the FCC should not continue to abdicate its responsibility for addressing the many problems facing the contemporary media environment.[58]

Reviving the fairness doctrine would involve telecommunications policy and, as such, is within the primary jurisdiction of the FCC rather than the courts. In this way, it is a policy-centred question, rather than a legal-centred one. Nor do I argue here that breathing life back into the fairness doctrine would solve any of the myriad challenges to democracy posed by disinformation discussed in this thesis.[59] However, the doctrine provides normative grounding in the USSC's First Amendment jurisprudence to contemplate approaches to media regulation based on the positive right of citizens to factual information in public discourse, as well as an express recognition that freedom of speech functions to facilitate 'the discovery and spread of political truth'.[60] There is significant value in considering these principles within the broader discourse in this area.

Neither European nor American courts appear to envisage a political era in which state actors abuse their positions by deliberately misinforming the public for political ends and actively endeavouring to undermine the important free speech role played by traditional media. Nor how advances in technology have made online speech such an effective tool in such efforts. This type of speech undermines the most important functions that free speech serves in a democracy and imperils democratic norms and institutions. While the express recognition of positive rights in the European framework make potential regulation in this area less controversial than in the United States, the fairness doctrine suggests that there may also be opportunities for more regulation in the American framework without contravening First Amendment principles. Indeed, such regulation may be imperative in order to protect such principles.

[57] ibid 3447.
[58] ibid.
[59] For example, for the purposes of FCC regulation, cable channels are treated differently from traditional broadcasters. See *FCC v Fox Television Stations, Inc*, 556 US 502 (2009).
[60] See *Consolidated Edison Co v Public Serv Comm'n*, 447 US 530, 534 (1980) (quoting Brandeis J's concurrence in *Whitney v California*, 274 US 357, 375 (1927)).

III. The Application of Human Rights Penality to Extreme Speech is Probably Counterproductive and Does More Harm than Good

An analysis of the role of harm and causation in the European framework, particularly in the context of online 'hate speech' and terrorist-related expression, reveals that the further that harm and causation are removed from legal frameworks, the greater the threat to freedom of expression. The recent proliferation of glorification-related offenses at the national and supranational levels in Europe examined in Chapter 4 illustrate this point. As Europe adopts increasingly broad definitions of terrorist offences, which are further and further attenuated from specific terrorist acts, the result is proscriptions on expression that bear no causal link to any risk of actual violence. The recent rise in prosecutions for such offences at the national level reflects the extent to which states are casting wide nets to capture speech that, while offensive and obnoxious, does not pose any legitimate threat to public safety.

Additionally, the discussion of human rights penality in Chapter 3, which cautions against the dangers of coercive overreach in the context of restrictions on 'hate speech', extends to other forms of extreme speech. As with glorification-related offences, criminal sanctions on 'hate speech' are extremely popular at the national level in Europe and are lacking in definitional rigour. Efforts at the national and supranational levels to commandeer intermediaries into policing their platforms based on such overly broad and vague offences raise serious concerns regarding the erosion of the protection of political speech in Europe and beyond.[61] Additionally, the absence of a developed methodology and rigorous scrutiny by the ECtHR in its Article 10 jurisprudence involving 'hate speech' and terrorist-related expression raises serious concerns regarding coercive overreach. This highlights the insufficient consideration in the European constitutional framework to the inherent ambiguity of human rights law as both limiting and requiring state coercion, with too much emphasis placed on the latter. This results in the ECtHR granting too much latitude and discretion to state efforts to place limitations on freedom of expression in the name of protecting vulnerable and marginalised groups from societal hatred and the public more broadly from acts of terrorist violence.

While the expansion of human rights penality is generally regarded as uncontroversial, important questions remain unresolved, including the most basic question of whether the

[61] As discussed infra in Part V, the regulation of digital intermediaries in the EU has important transnational implications.

protection of human rights requires criminal accountability.[62] In addition to such questions, there is the practical question of whether human rights penality is an effective tool for ameliorating the societal problems to which proscriptions on extreme speech are directed. As discussed in Chapter 3, criminal proscriptions on 'hate speech' are directed to protecting the targets of such expression from abuse and reducing levels of societal hate. Research concerning recent trials of European politicians for violating national 'hate speech' legislation suggest that the success of human rights penality in this context is questionable, particularly in cases involving politicians.

Over the past several years, several anti-immigration politicians in Europe have faced prosecution for violating criminal proscriptions on 'hate speech' at the national level.[63] While few systematic analyses have been conducted of the effect of prosecutions and/or convictions on voters' behaviour, examples such as Jean Marie Le Pen and Geert Wilders suggest that 'hate speech' prosecutions may enhance rather than decrease the popularity of populist politicians.[64] The results of a 2020 study of 'hate speech' trials in The Netherlands, Germany, France and Belgium involving an aggregative-level analysis suggest that 'hate speech' prosecutions of prominent populist political figures do not cause any political harm to these figures in terms of an electoral penalty.[65] Indeed, none of the cases studied found negative correlations between news visibility of a hate speech prosecution and electoral support. This suggests that exposure to news about criminal proceedings against a politician not only does not erode electoral support for the associated political party but may actually be advantageous in some cases by providing the party with free publicity and a prominent public platform.[66]

In stark contrast to the emerging European trend of expanding traditional notions of harm and causation within criminal law frameworks in the context of extreme speech, the United States situates extreme speech firmly within existing frameworks. For example, the *Brandenburg* test does not permit a state to proscribe advocacy of the use of force or law violation except where such advocacy is directed to both inciting or producing imminent lawless action *and* is likely

[62] Mattia Pinto, 'Historical Trends of Human Rights Gone Criminal (2020) 42(4) HRQ 729, 731 - 732.
[63] These include Jean Marie Le Pen in France and Geert Wilders in the Netherlands. See Laura Jacobs and Joost Van Spanje, 'Prosecuted, yet popular? Hate Speech Prosecution of anti-immigration politicians in the news and electoral support' (2020) 18 Comp Eur Polit 899.
[64] Heli Askola, 'Taking the Bait? Lessons from a Hate Speech Prosecution' (2014) 30(1) CJLS 51, 52.
[65] See Jacobs and Van Spanje (n 63).
[66] ibid.

to incite such action.[67] As discussed in Chapter 4 in the context of terrorist-related expression, a party attempting to hold a digital intermediary liable for third party content on its platform must establish that such expression was the *proximate* cause of injuries sustained in a subsequent act of terrorism. Generalised links between the use of social media by terrorist groups and terrorist acts are insufficient, as are conclusory allegations regarding the harm that results from terrorist-related content on intermediary platforms.

The American approach is consistent with the international framework concerning extreme speech and violence, which recommends that proscriptions on hateful expression focus exclusively on the advocacy of discriminatory hatred that constitutes incitement to hostility, discrimination, or violence - rather than the advocacy of hatred without regard to its tendency to incite action by the audience - and should require a likelihood of imminent harm and a reasonable probability of discrimination, hostility, or violence occurring as a direct consequence of such incitement.[68] The UN also advises that only incitement to discrimination, hostility, or violence that meets all the Rabat Plan criteria should be criminalised, and that less severe forms of incitement or 'hate speech' warrant civil or administrative law or public policy responses.[69] The international framework is particularly instructive in this context as criminal law responses do not appear particularly effective at reducing societal hate and violence, and a preoccupation with criminality shifts the focus of public discourse to individual private actors - those who engage in 'hate speech' and terrorist-related expression - rather than the broader systemic social and cultural problems that fuel societal hate and terrorist violence.

IV. The ECtHR's Methodology in the Context of Extreme Speech Would Benefit from Greater Conceptual Clarity and Coherence

In interrogating the approach of the ECtHR in extreme speech cases, it is important to bear in mind its status as an international court. As discussed in the Introduction, the ECtHR belongs to a different type of legal order than the USSC. Unlike the latter, the ECtHR relies in large part on the consent of Member States with respect to its legitimacy and authority, and is guided by the principle of subsidiarity and the doctrine of the margin of appreciation. However, this

[67] See *Brandenburg v Ohio*, 395 US 444, 447 (1969).
[68] See discussion of the Rabat Plan in Chapter 3, pp 92 - 93.
[69] 'UN Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences' (September 2020) <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf> accessed 4 December 2020.

does not insulate the Court from criticism, and the absence of a coherent methodology and conceptually precise terminology in the Court's approach to extreme speech merits demanding scrutiny.

Chapters 1, 3, and 4 offer critical analysis of the ECtHR's methodology in extreme speech cases. As noted in Chapter 1, the ECtHR's application of the margin of appreciation, in particular, is the subject of widespread criticism from both scholars and judges for a lack of clarity and consistent application.[70] Ian Cram offers a persuasive argument that the ECtHR's application of the margin of appreciation doctrine has resulted in an excessive degree of discretion granted to national courts in their regulation of offensive speech and the conferral on Member States of wide and vaguely defined powers to prescribe the manner in which ideas and opinions are expressed.[71] Cram is particularly concerned with the ECtHR's treatment of offensive speech concerning religious subjects and observes that in such cases the Court 'has readily backed away from interfering with national authorities' interference with expression'.[72] This is particularly apparent in cases involving expression concerning Islamic links to terrorist activity, such as statements in support of a Caliphate.[73] Cram argues that this type of speech falls within the common understanding of 'political speech' and, as a result, the Court should apply rigorous scrutiny to national authorities' restrictions on such expression. The absence of such scrutiny, Cram warns, may lead to a chilling effect on freedom of expression and stifling of public debate concerning religious fundamentalism and terrorism.[74]

Cram's observations regarding the ECtHR's relaxed scrutiny in the context of terrorist-related expression augment the discussion in Chapter 4, which highlights that while the Court emphasises that there is little scope under Article 10 for proscriptions on political speech or on debate of questions of public interest where the views expressed do not comprise incitement to violence, it does not elaborate on how expression that justifies the commission of terrorist offences constitutes incitement to violence. Moreover, it has found no violation of Article 10 in cases involving expression that falls well short of traditional conceptions of incitement.[75]

---

[70] See Chapter 1, p 28.
[71] Ian Cram, 'The Danish Cartoons, Offensive Expression, and Democratic Legitimacy' in Ivan Hare and James Weinstein (eds), *Extreme Speech and Democracy'* (OUP 2009).
[72] ibid 15.
[73] ibid 324. See *Refah Partisi (the Welfare Party) and others v Turkey,* App nos 41340/98, 41342/98, 41343/98 and 41344/98 (ECtHR, 13 February 2003).
[74] ibid.
[75] See discussion of *Leroy v France,* App No 36109/03 (ECtHR, 10 February 2008) in Chapter 4, p 142.

Doctrinal uncertainty results from the absence of a clear delineation in the Court's methodology regarding political speech involving terrorism that warrants protection under Article 10 and political speech involving terrorism that falls outside of such protection.

The Court's 'hate speech' jurisprudence also suffers from a lack of conceptual clarity and underdeveloped methodology. Chapter 3 examines this issue through an in depth discussion of the 2020 case of *Lilliendahl v. Iceland*, which illustrates that the Court's decision not to define 'hate speech' has led to a lack of clarity concerning what types of hateful expression constitute 'hate speech' and how to distinguish between the 'gravest' and 'less grave' forms within this category.[76] This conclusion is consistent with the Future of Free Speech project's[77] (the Project) analysis of sixty ECtHR and European Commission on Human Rights 'hate speech' cases between 1979 and 2020.[78] The objective of the Project's analysis was to evaluate the application of the Court's holding that Article 10 is applicable to information ideas that shock, offend, or disturb in 'hate speech' cases.[79] Based on this extensive analysis, the Project concluded that in 'hate speech' cases the threshold for the Court finding that an interference did not violate Article 10 is low and its application of Article 10 is inconsistent. It notes that while the ECtHR acknowledges that it is necessary in certain democratic societies to limit some type of speech, it 'has yet to provide a coherent legal and/or normative extrapolation of when/where/how these limitations can or should occur'.[80] This, the Project argues, 'has resulted in anomalies vis-à-vis the treatment of similar cases'.[81]

The Project recommends that a 'proper definitional framework' is imperative for the application of a 'coherent set of thresholds'.[82] It is argued here that such a framework should provide the necessary clarity regarding the Court's methodology with respect to the application of the margin of appreciation in extreme speech cases, identify the circumstances in which the application of Article 17 is appropriate in such cases, and articulate the proper boundary between political speech that is protected under Article 10 and political speech that falls outside

[76] See Chapter 3, p 106.
[77] The Future of Free Speech Project is run by Justitia in collaboration with Columbia University's Global Freedom of Expression project and Aarus University's Department of Political Science.
[78] Jacob Mchangama and Natalie Alkiviadou, '"Hate Speech" Jurisprudence of the ECtHR through a Qualitative and Quantitative Lens' (*ECHR Blog,* 3 November 2020) <https://www.echrblog.com/2020/11/guest-post-hate-speech-jurisprudence-of.html> accessed 2 December 2020.
[79] See *Handyside v United Kingdom* (1976) ECHR 5.
[80] Mchangama and Alkiviadou (n 78).
[81] ibid.
[82] ibid.

the scope of the ECHR. This last point is particularly important because the lack of conceptual clarity concerning the Court's application of the margin of appreciation in these types of cases creates a danger that unpopular political expression and dissent on important matters of public concern and debate will be excised from public discourse. Given the popularity of criminal sanctions in Member States that target online expression, the current approach of the ECtHR presents a significant risk to the adequate protection of freedom of expression in the digital age.

V. The Compulsory Regulation of Digital Intermediaries Raises Transnational Free Speech Concerns and Challenges

The examination of the EU's shift to a compulsory framework for the regulation of digital intermediaries in Chapter 4 highlights the transnational free speech implications of efforts to regulate online terrorist-related content. Such implications form part of a larger discourse involving increasingly aggressive efforts by states and supranational bodies to regulate the content of intermediary platforms, in particular, in the context of extreme speech. In a surprising development given the expansive protection of freedom of expression and the broad immunity afforded to digital intermediaries under the CDA, these efforts include those of state legislatures in the United States. For example, in June 2021, the Republican Governor of Florida, Ron DeSantis, signed a law that makes it illegal for digital intermediaries to, among other things, bar a candidate for state office for more than 14 days.[83] In defending the law, Governor DeSantis claimed that the legislation protected Floridians from online social media censorship.[84] In his order granting a preliminary injunction enjoining enforcement of parts of the legislation based on the First Amendment, a federal judge acknowledged that '[w]here social media fit in traditional First Amendment jurisprudence is not settled'.[85] However, he went on to opine that the state's assertion that it was 'on the side of the First Amendment' was 'wholly at odds with accepted constitutional principles':

> The First Amendment says "Congress" shall make no law abridging the freedom of speech or of the press. The Fourteenth Amendment extended this prohibition

[83] See David McCabe, 'Florida, in a First, Will Fine Social Media Companies That Bar Candidates' *The New York Times* (24 May 2021) <https://www.nytimes.com/2021/05/24/technology/florida-twitter-facebook-ban-politicians.html> accessed 1 July 2021.
[84] 'Gov. DeSantis announces legislation to crack down on big tech, online censorship' *ABC News* (2 February 2021) <https://www.wtxl.com/news/local-news/gov-desantis-announces-legislation-to-crack-down-on-big-tech-online-censorship> accessed 10 May 2021.
[85] *Netchoice, LLC v Moody*, Case No. 4:21cv220-RH-MAF (Preliminary Injunction, 30 June 2021) Doc 113, 17.

> to state and local governments. The First Amendment does not restrict the rights of private entities not performing traditional, exclusive public functions…So whatever else may be said of the providers' actions, they do not violate the First Amendment.[86]

Republican lawmakers in several states have recently introduced similar bills that would allow for civil lawsuits against platforms for what they characterise as the 'censorship' of posts.[87] This relates to the debunked claim that social media is biased against conservative voices.[88] These examples highlight that aggressive efforts to regulate digital intermediaries extend beyond terrorist-related expression and that such efforts extend to jurisdictions like the United States with expansive protection on speech in public discourse and limited conceptions of horizontality.

Debates over intermediary liability in the United States implicate horizontality with respect to the application of the state action doctrine. As discussed in Chapter 1, the second strand of the state action doctrine triggers the application of constitutional protections to acts from private actors in circumstances in which the government becomes entangled with private parties by, for example, compelling the private entity to take a particular action.[89] In the context of state efforts to force intermediaries to host particular types of expression on their platforms, such efforts may implicate the entanglement strand of the state action doctrine, thus potentially triggering the application of the First Amendment to private expression on social media platforms. However, as emphasised throughout this thesis, digital intermediaries, as private actors, have First Amendment rights. At present, efforts to regulate digital intermediaries are stymied by the USSC's interpretation of the scope of the First Amendment as well as the CDA immunity for digital intermediaries.[90] However, as states take increasing interest in regulating the content of social media platforms, the greater the risk that a constitutional quagmire will develop. In this context, it may be instructive to look to Europe to both frame and enrich a discussion of horizontality in the United States.

---

[86] ibid.
[87] Anthony Izaguirre, 'GOP pushes bills to allows social media 'censorship' lawsuits' *AP News* (7 March 2021) <https://apnews.com/article/donald-trump-legislature-media-lawsuits-social-media-848c0189ff498377fbfde3f6f5678397> accessed 10 March 2021.
[88] See Paul M. Barrett and J. Grant Sims, 'False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives' (*NYU Stern Center for Business and Human Rights*, February 2021).
[89] See Chapter 1, p 48.
[90] For an overview of the CDA, see Chapter 4, pp 144 - 145.

While the discourse concerning the free speech implications of aggressive efforts to regulate digital intermediaries is constantly evolving due to changes in regulatory frameworks as well as new developments in content moderation policies, it is worthwhile to examine the current landscape regarding the regulation of intermediary platforms. Daphne Keller, Director of Intermediary Liability at the Center for Internet and Society at Stanford Law School, warns of the dangers to the free speech rights of users when states delegate interpretation and enforcement of speech laws to private actors. Drawing on data from studies concerning intermediary liability in the sphere of copyright law, Keller observes that '[t]wenty years of experience with these laws in the United States and elsewhere tells us that when platforms face legal risk for user speech, they routinely err on the side of caution and take it down'.[91] She further observes that legally mandated notice and takedown systems invariably favour accusers as, in the face of potential liability, the easiest and cheapest route available to intermediaries is to simply remove the objectionable content.[92] This avoids both the risk of contravening the law in a given jurisdiction and the expense of paying individuals to assess the validity of claims.[93] In short, while intermediaries appear to exercise their own discretion in removing content, their decisions are often profoundly influenced by governments.[94] As such, intermediaries 'anticipatory obedience spares governments the need to enact actual laws - and deprives affected users the opportunity to challenge them in court'.[95]

With respect to Europe, in particular, Keller asserts that the EU 'is very much in the driver's seat in regulating major platforms'.[96] Specifically, whatever the EU compels major platforms to do, they are likely to do everywhere, potentially 'voluntarily' by changing their global Terms of Service (TOS).[97] In discussing a draft version of TERREG, Keller emphasised that the contemplated compulsory framework significantly increased the incentives for intermediaries

[91] Daphne Keller, 'Internet Platforms: Observations on Speech, Danger, and Money,' Hoover Institution's Aegis Paper Series, No. 1807 (13 June 2018).
[92] ibid.
[93] ibid.
[94] Daphne Keller, 'Who Do You Sue? State And Platform Hybrid Power Over Online Speech', A Hoover Institution Essay, Aegis Series Paper No. 1902 (29 January 2019).
[95] ibid 2. See also Ben Wagner, "Free Expression? Dominant Information Intermediaries as Arbiters of Internet Speech," in Martin Moore and Damian Tambini (eds), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (OUP 2018) 225 ('[a]n implicit bargain between platforms and regulators results in government's advocating for self-regulatory measures that would not otherwise be legal if they were enshrined in law'.).
[96] Daphne Keller and Joan Barata, 'Regulating Online Terrorist Content: A Discussion with Stanford CIS experts about New EU Proposals' (*Stanford Law School*, 25 April 2019) <https://law.stanford.edu/2019/04/25/regulating-online-terrorist-content-a-discussion-with-stanford-cis-directors-about-new-eu-proposals/> accessed 4 November 2019.
[97] ibid.

to prohibit content using private TOS rather than national law, and to remove more material than national laws require.[98] At the same time, compulsory regulation increases the power of any one EU Member State to censor information that is legal in other parts of the EU.[99] For example, when two EU Member States disagree on whether a particular piece of content promotes or glorifies terrorism, TERREG incentivises intermediaries to use their TOS to accommodate the authority that wants to the content removed, and to apply the strictest legal standard to users everywhere.[100] That digital intermediaries will censor material on a global scale based on the countries with the most rigorous censorship laws is a key concern of those worried about the transnational free speech implications of regional efforts to regulate digital intermediaries. For example, American law professor Danielle Citron observes that '[a]s companies alter speech rules and speech operations in a wholesale way (rather than retail via country), then the strictest regime prevails', which 'is a considerable threat to free expression'.[101]

While acknowledging that data about removal practices is limited, Keller identifies research out of the University of California, Berkeley and Columbia University that documents significant over-removal of content under the Digital Millennium Copyright Act, notwithstanding the fact that the legislation contains provisions aimed at preventing this very problem.[102] Moreover, over-removal resulting from aggressive efforts to counter extremism is now considered 'commonplace'; examples include YouTube removing videos of Syrian atrocities posted by a UH Human Rights organisation and Facebook deleting the page of a Chechen pro-independence group opposed to terrorism and posts documenting Rohingya ethnic cleansing in Myanmar, reportedly because it had classified Rohingya organisations as dangerous militant groups.[103]

---

[98] Daphne Keller, 'The EU's Terrorist Content Regulation: Expanding The Rule of Platform Terms of Service and Exporting Expression Restrictions From the EU's Most Conservative Member States' (*Center for Internet and Society,* 25 March 2019) <http://cyberlaw.stanford.edu/blog/2019/03/eus-terrorist-content-regulation-expanding-rule-platform-terms-service-and-exporting> accessed 20 January 2019.
[99] ibid.
[100] ibid.
[101] David L. Hudson Jr., 'Free speech or censorship? Social media litigation is a hot legal battleground' (*ABA Journal* 1 April 2019) <https://www.abajournal.com/magazine/article/social-clashes-digital-free-speech> accessed 11 July 2021.
[102] Keller, 'Internet Platforms' (n 106) 3.
[103] ibid. See also Diana Lee, 'Germany's NetzDG and the Threat to Online Free Speech (*Media Freedom & Information Access Clinic at Yale Law School*, 10 October 2017) <https://law.yale.edu/mfia/case-disclosed/germanys-netzdg-and-threat-online-free-speech> accessed 11 March 2020 (emphasising that Germany's NetzDG law underscores the costs of imposing intermediary liability for unlawful content online by encouraging social medial platforms to over-police speech, thereby chilling legitimate expression and discourse on significant public debate forums).

The work of Keller and others raise issues of concern to those interested in the potential free speech implications resulting from the emerging European trend of placing increasingly onerous obligations on intermediaries to regulate online content. Recent additions to this trend are the EU's Digital Services Act package and the UK's Draft Online Safety Bill (Draft Safety Bill). Introduced by the European Commission in December of 2020, the former includes two proposed pieces of EU legislation - the Digital Services Act (DSA) and the Digital Market Act. One of the primary aims of the DSA[104] is to consolidate various pieces of EU legislation and voluntary regulatory practices that relate to harmful or illegal online content, and to harmonise the rules concerning the provision of digital services across the EU.[105] In commenting on the DSA proposal, ARTICLE 19 notes that it goes further than consolidation and harmonisation by seeking to make large digital intermediaries, such as Twitter and Facebook, answerable to public authorities by way of new transparency and due diligence obligations.[106] One of ARTICLE 19's primary concerns is that the DSA 'effectively empowers hosting providers to make decisions about the legality of content upon receipt of a substantiated notice of alleged illegality. Since substantiated notices constitute actual knowledge…for the purposes of the hosting immunity under Article 5, hosting providers have a strong incentive to remove content upon notice'.[107] ARTICLE 19 further notes that while the DSA proposal includes safeguards for the protection of freedom of expression, that 'in practice, it will be easier for [intermediaries] to remove content to avoid any liability risk'.[108]

In May of 2021, the UK government issued the Draft Safety Bill, a 145 page document that sets out a new regulatory framework to address 'harmful' content online that includes imposing duties of care on digital intermediaries to protect users from illegal content from other users as well as affirmative safeguarding measure.[109] While a detailed analysis of the Draft Safety Bill is outside the scope of this thesis, it is worth highlighting that numerous civil society organisations, including ARTICLE 19, Index on Censorship, and the Electronic Frontier

[104] See 'The Digital Services Act package' (*European Commission*) <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> accessed 15 July 2021.
[105] ibid.
[106] 'At a glance: Does the EU Digital Services Act protect freedom of expression?' (*ARTICLE 19,* 11 February 2021) <https://www.article19.org/resources/does-the-digital-services-act-protect-freedom-of-expression/> accessed 11 July 2021.
[107] ibid.
[108] ibid.
[109] 'Draft Online Safety Bill' (*Department for Digital, Culture, Media & Sport*, May 2021) ISBN 978-1-5286-2563-0. The Draft Safety Bill follows the Online Harms White Paper published by the UK government in 2019.

Foundation have expressed significant concerns regarding its potential impact on freedom of expression. These include concerns over the proposed regime for sanctions, which imposes significant fines to criminal liability in some circumstances as well as vague and imprecise definitions of key terms including 'legal but harmful' content.[110]

Regarding the DSA, discussions in the European counsel to find a common position are ongoing.[111] While EU Member States appear generally supportive of the proposal, discussions concerning enforcement, content moderation, and other issues are also ongoing.[112] The Joint Committee on the Draft Online Safety Bill, established by the House of Lords and the House of Commons, will submit its report on the proposed legislation in December of 2021. Thus, at the time of the writing it is impossible to know what the DSA and Draft Safety Bill will look like if and when each piece of proposed legislation becomes law. With that said, this proposed legislation illustrates a broader trend in Europe of replacing voluntary frameworks with compulsory ones that place onerous obligations on intermediaries to moderate and remove content. The impact of such efforts on freedom of expression, in Europe and beyond, will be the subject of future research.

Conclusion

This chapter offers some reflections and recommendations that emerge from the comparative analysis of the European and American approaches to the regulation of 'hate speech', terrorist-related expression, and disinformation from state actors undertaken in prior chapters in light of the primary objective of this thesis – to offer insights into the broader debate concerning the appropriate limits on freedom of expression in liberal democracies in the digital age. The discussions in this chapter highlight the extent to which the digital age has unsettled longstanding normative and doctrinal elements of European and American constitutional frameworks. As technology continues to transform the ways individuals and state actors participate in and contribute to public discourse, so too will debates over how to adequately protect freedom of expression while combatting the challenges posed by extreme speech. Final

[110] See 'UK: Draft Online Safety Bill poses serious risk to free expression' (*ARTICLE 19*, 26 July 2021) <https://www.article19.org/resources/uk-draft-online-safety-bill-poses-serious-risk-to-free-expression/> accessed 27 July 2021.
[111] 'Legislative Train 06.2021, Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC / After 2020-9' (*European Parliament*, 2021) <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act> accessed 10 July 2021.
[112] ibid.

reflections and additional areas for future research are discussed in the Conclusion to this thesis.

Conclusion

As emphasised in the Introduction and illustrated by the examinations undertaken in Parts I and II, while the goal of free speech doctrine is straightforward - to restrain the government from suppressing speech that democracies must permit while allowing it to proscribe speech that causes harm that it may legitimately prevent - accomplishing this goal is anything but.[113] This is particularly true in the digital age as advancements in technology open new pathways for communicating in public discourse that present new challenges and opportunities for speech regulation.

This thesis illustrates that a CCL inquiry of the European and American approaches to 'hate speech', terrorist-related expression and disinformation from state actors in one volume is particularly instructive in addressing the broader question of the appropriate limits on freedom of expression in liberal democracies in the digital age. For example, an inquiry of this nature, which rigorously examines the nuances of relevant doctrines and philosophical principles and delves into both inter and intra-jurisdictional debates regarding the proper scope and application of free speech rights, captures the nuances and contested elements within each approach. Too often, the comparative discourse surrounding the regulation of extreme expression in Europe and the United States is framed in high level terms, notwithstanding the fact that examination of the nuances and contested elements in each approach create opportunities for the most interesting and useful insights to emerge.

Consideration of different types of extreme speech in one volume also illuminates the range of harms at issue and the distinct and sometimes overlapping dangers - both to individuals, groups, and the public - as well as to free speech flowing therefrom. For example, while at present there is no strong empirical evidence linking terrorist-related expression with subsequent terrorist acts, there is abundant evidence that glorification-related offences imperil free speech and permit states to restrict unpopular expression that poses no danger to safety and security. Thus, this thesis is critical of the emerging trend in Europe of proscribing expression that lies well outside of traditional conceptions of incitement based on highly attenuated or non-existent

[113] See Introduction, p 1 (James Weinstein, *Hate Speech, Pornography, and Radical Attacks on Free Speech Doctrine* (Routledge 1999) 11).

empirical evidence as well as the lack of clarity and coherence in the European approach to regulating terrorist-related expression, and argues that the speech protective approach of the United States appears preferable in this area. Moreover, while terrorism represents a small risk to Western democracies, the threat of terrorist violence is commonly used as a justification for sweeping restrictions on speech that endanger the protection of unpopular political expression and dissent.

In contrast, online disinformation, particularly from domestic state actors, destabilises the liberal democratic order and, in so doing, impedes the ability of states to adequately protect vulnerable and historically marginalised groups from societal hatred and to safeguard the broader public from acts of terrorist violence.[114] The examination of domestic disinformation campaigns in Chapter 5, particularly the campaign directed to undermining the integrity of the 2020 presidential election in the United States, support this conclusion. This form of disinformation also *foments* the type of societal hatred and violence that regulatory efforts directed to 'hate speech' and terrorist-related expression are aimed at ameliorating. For this reason, it is reasonable to conclude that disinformation from state actors is the most dangerous form of extreme speech in the digital age. The emerging European approach to the dangers of domestic disinformation from state actors in the digital age is preferable to the American approach to the extent that the latter appears ill-equipped to adequately acknowledge, let alone address, such dangers. However, the undeveloped methodology of the ECtHR in the area of extreme speech as well as the widespread acceptance of content and viewpoint-based proscriptions on expression at the national and supranational levels in Europe are cause for concern in the context of emerging regulatory frameworks.

As we consider how to tackle the challenges posed by disinformation from state, recent efforts in Europe to regulate other types of extreme speech warrant attention. For example, the application of human rights penality is likely not the answer as the use of criminal sanctions are probably counter-productive and may serve to increase public support for state actors who engage in disinformation campaigns. Additionally, the longstanding problems with respect to regulatory frameworks that target 'hate speech' caution against the use of conceptually

[114] See, e.g., Asif Efrat and others, 'Report on the Relationship Between Terrorist Threats and Governance Condition in the European Union' (*RECONNECT,* 29 June 2021) <https://reconnect-europe.eu/wp-content/uploads/2021/06/D11.3.pdf)> accessed 1 July 2020, 15: 'Processes of democratic backsliding or outright "autocratization" bring countries into the "danger zone" of substantially increased probability of relatively high numbers of terrorist attacks and casualties.'

imprecise (or nonexistent) definitions that provide too much space for proscriptions on politically unpopular expression that may reasonably be classified as political speech. Similarly, an examination of the potential transnational implications of aggressive approaches to the regulation of digital intermediaries in the area of terrorist-related expression in Europe is instructive to the broader global discourse regarding intermediary liability for the extreme speech of users.

Additionally, the ways in which the digital age has transformed public discourse by fundamentally altering the ways in which individual communicate with one another and how governments communicate with the public should guide evaluations of whether certain doctrines warrant re-examining - or abandoning altogether - and how existing principles relating to the application of human rights to relationships between private actors are implicated. Again, a comparative approach is useful as the horizontality discourse in Europe is more developed and robust than in the United States and, as a result, may provide a useful framework for discussions regarding the proper scope of the state action doctrine in the digital age. Another theme that emerges from this CCL inquiry is that regulations on extreme speech reflect decisions by states and supranational bodies regarding what types of expression are permissible in public discourse, who may legitimately advocate for the use of violence and for what ends, and the types of harm that justify restrictions on expression. This thesis illuminates the political nature of such decisions and how they become significantly more complex when state actors are recognised as important speech actors in public discourse.

The examinations undertaken in this thesis further reveal that the relevant question is not simply what approach is better suited to the challenges of the digital age, as these challenges are nuanced and complex and, thus, so too must be any proposed solutions. Moreover, this analysis illuminates that certain elements of each approach are better suited to addressing the unique challenges that arise from particular types of extreme speech. For example, while the emphasis on human rights penality in the context of 'hate speech' and terrorist-related expression in Europe is most likely not an effective response to the societal challenges posed by societal hatred and terrorist violence, the European approach to disinformation, which considers the unique challenges posed by disinformation from domestic state actors, may be better equipped to address the danger posed by this type of extreme speech. Thus, simply advocating for one approach over the other fails to account for important distinctions both in

terms of the nuances within each approach and the fact that certain approaches may be better suited to tackling the unique challenges flowing from different types of speech. While the comparative discourse should be framed by such distinctions, from which the most useful insights and lessons emerge, it is too often framed in generalities.

Ultimately, these examinations suggest that the American approach to disinformation is most in need of urgent re-examination and reform in the digital age. This is not to suggest that the deficiencies in the European approach identified and analysed in this thesis, particularly with respect to 'hate speech' and terrorist-related expression, are unimportant but, rather, that the examination undertaken in Chapter 5 reveals that the harm flowing from domestic disinformation from state actors poses an existential threat to contemporary American democracy. Unlike Europe, which has safeguards against democratic safeguarding built into its norms and institutions, an unwavering belief in the stability and inevitability of democracy predominate the American approach. This may be understood, at least in part, from the different historical contexts in which the United States Constitution and the ECHR emerged; the former out of a newly formed democracy at the end of the eighteenth century, and the latter out of a post-World War II Europe that had recently confronted fascism and the dangers of totalitarianism. Confidence in its version of democracy, coupled with American exceptionalism, has resulted in an unwillingness to meaningfully grapple with the challenges and dangers posed by domestic disinformation from state actors in the digital age.

Another theme that emerges from this comparative inquiry is that while the digital age has raised new challenges and dangers relating to extreme speech, the underlying societal problems that extreme speech exposes are not new. Indeed, societies have been grappling with hatred against marginalised and vulnerable communities, terrorist violence, and disinformation for centuries. Thus, technology did not create these problems and cannot, by itself, be expected to fix them. Nor can state and supranational efforts to regulate such technology. Accordingly, while existing normative and doctrinal frameworks warrant re-examining to ensure that they meet the challenges of the digital age while providing adequate safeguards for freedom of expression, we must not lose sight of the fact that the problems we are confronting are much older and firmly entrenched. Accordingly, solving these problems will require more than efforts to regulate online platforms and more tools than the law has at its disposal.

As emphasised in the Introduction, this thesis is current as of July of 2021. Subsequent events and developments will be the subject of future research projects and may alter the recommendations and conclusions reached in this volume. Given the emergence of aggressive efforts to regulate the content of intermediary platforms and the transnational debates relating thereto, intermediary liability will undoubtedly be one area of future research. American judges appear willing to contemplate a shift in the way in which digital intermediaries are regulated in the American constitutional framework. Justice Clarence Thomas offered an example of such willingness in his concurrence in the *Knight* case.[115] When exactly this time will come remains to be seen. However, immediate and rigorous scrutiny of any such efforts will be warranted.

Another area of future research involves whether and to what extent international law has a part to play in the discourse concerning the proper scope of freedom of expression in the digital age. Chatham House recently noted that with its 'careful calibrations designed to protect individuals from abuse of power by authority', international law provides a useful normative framework for responses to extreme speech.[116] As discussed in Chapter 3, the Rabat Plan provides a useful rubric for evaluating efforts to regulate 'hate speech'. Given globalisation and the transnational implications of national and supranational efforts to regulate online speech, research regarding the benefits and potential drawbacks of an international normative framework for the regulation of extreme speech is worthwhile to the overarching question that drives this research - the appropriate limits on freedom of expression in liberal democracies in the digital age.

[115] See Chapter 6, pp 201 - 202.
[116] Kate Jones, 'Online Disinformation and Political Discourse: Applying a Human Rights Framework' (*Chatham House International Law Program*, 6 November 2019) 2 <https://www.chathamhouse.org/2019/11/online-disinformation-and-political-discourse-applying-human-rights-framework> accessed 3 July 2021.

Bibliography

ABC News, 'Gov. DeSantis announces legislation to crack down on big tech, online censorship' *ABC News* (2 February 2021) <https://www.wtxl.com/news/local-news/gov-desantis-announces-legislation-to-crack-down-on-big-tech-online-censorship> accessed 10 May 2021

Agha P (ed), *Human Rights Between Law and Politics: The Margin of Appreciation in Post-National Contexts* (Hart Publishing 2017)

Alkiviadou N and Mchangama J, 'Hate Speech and the European Court of Human Rights: Whatever Happened to the Right to Offend, Shock, or Disturb?' (2021) 00 HRL Rev 1

— — '"Hate Speech" Jurisprudence of the ECtHR through a Qualitative and Quantitative Lens' (*ECHR Blog,* 3 November 2020) <https://www.echrblog.com/2020/11/guest-post-hate-speech-jurisprudence-of.html> accessed 2 December 2020

Allen MC, 'Parliamentary immunity in the members states of the European Union and in the European Parliament' (*Legal Affairs (Series W-4),* February 1993)

Althuis J, Granelli F and Colley T, 'Disinformation's Societal Impact: Britain, Covid, And Beyond' (2020) 8 Defence Strategic Communications 89

Amnesty International, 'Written contribution the thematic discussion on Racist Hate Speech and Freedom of Opinion and Expression organized by the United Nations Committee on Elimination of Racial Discrimination' (*Amnesty International*, 28 August 2012) <https://www.amnesty.org/download/Documents/24000/ior420022012en.pdf> accessed 28 September 2020

— — 'Statement of Amnesty International', UN High Level Conference of Heads of Counter-Terrorism Agencies of Member States (*Amnesty International*, 29 June 2018) <https://www.un.org/counterterrorism/ctitf/sites/www.un.org.counterterrorism.ctitf/files/S4-Amnesty-International.pdf> accessed 12 February 2019

— — 'Tweet...If You Dare: How Counter-Terrorism Laws Restrict Freedom of Expression in Spain' (*Amnesty International*, 2018) <https://www.amnesty.org/download/Documents/EUR4179242018ENGLISH.PDF> accessed 15 April 2019

Aral S, Roy D and Vosoughi S, 'The Spread of True and False News Online' (2018) 359(6380) Science 1146

Arendt H, *Crises of the Republic: Lying in Politics; Civil Disobedience; on Violence; Thoughts on Politics and Revolution* (1st edn, Mariner Books 1972)

Argomaniz J, Bures O and Kaunert C, 'A Decade of EU Counter-Terrorism and Intelligence: A Critical Assessment' (2015) 30(2-3) Intelligence and National Security 191

ARTICL19, ‘Prohibiting incitement to discrimination, hostility or violence,’ Policy Brief (*ARTICLE* 19, December 2012) <https://www.refworld.org/pdfid/50bf56ee2.pdf> accessed 20 September 2020

— — ‘“Hate Speech” Explained: A Toolkit (2015 Edition)’ (*ARTICLE 19*, 2015) <https://www.article19.org/resources/hate-speech-explained-a-toolkit/> accessed 4 November 2020.

— — ‘Responding to ‘Hate Speech’: Comparative Overview of Six EU countries’ (*ARTICLE 19,* 2018) <https://www.article19.org/resources/responding-hate-speech-comparative-overview-six-eu-countries/> accessed 20 September 2020

— — ‘Germany: Responding to “hate speech”’ (*ARTICLE 19 Country Report*, 2018) <https://www.article19.org/wp-content/uploads/2018/07/Germany-Responding-to-%E2%80%98hate-speech%E2%80%99-v3-WEB.pdf> accessed 3 November 2020,

— — ‘Spain: Speech related offences of the Penal Code’ (*ARTICLE 19*, March 2020) <https://www.article19.org/wp-content/uploads/2020/03/Spain-Penal-Code-analysis-March-2020-Final.pdf> accessed 4 August 2020.

— — ‘At a glance: Does the EU Digital Services Act protect freedom of expression?’ (*ARTICLE 19,* 11 February 2021) <https://www.article19.org/resources/does-the-digital-services-act-protect-freedom-of-expression/> accessed 11 July 2021.

— — ‘UK: Draft Online Safety Bill poses serious risk to free expression’ (*ARTICLE 19*, 26 July 2021) <https://www.article19.org/resources/uk-draft-online-safety-bill-poses-serious-risk-to-free-expression/> accessed 27 July 2021.

Askola H, ‘Taking the Bait? Lessons from a Hate Speech Prosecution’ (2014) 30(1) CJLS 51

Baade B, ‘Don’t Call A Spade A Shovel: Crucial Subtleties in the Definition of Fake News and Disinformation’ (*Verfassungsblog,* 14 April 2020) <https://verfassungsblog.de/dont-call-a-spade-a-shovel/> accessed 20 May 2021

Baer S, ‘Dignity, Liberty, Equality: A Fundamental Rights Triangle of Constitutionalism’ (2009) 59(4) U Toronto L J 417

Bail CA and others, ‘Exposure to opposing views on social media can increase political polarization’ (2018) 115(37) PNAS 9216

Bailey H, Samantha Bradshaw and Philip N Howard, ‘Industrialised Disinformation: 2020 Global Inventory of Organized Social Media Manipulation’ (*Computational Propaganda Project at the Oxford Internet Institute*, 2020) <https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/> accessed 28 April 2020

van Ballegooij W and Bakowski P, ‘The Cost of Non-Europe in the fight against terrorism’ (*European Parliamentary Research Service*, May 2018) <https://www.europarl.europa.eu/RegData/etudes/STUD/2018/621817/EPRS_STU(

Barak A, *Human Dignity: The Constitutional Value and the Constitutional Right* (CUP 2015)

Barata J, ‘New EU Proposal on the Prevention of Terrorist Content Online: An Important Mutation of the E-Commerce Intermediaries’ Regime (White Paper)’ (2018) (*The Center for Internet and Society at Stanford Law School*, 12 October 2018) <http://cyberlaw.stanford.edu/publications/new-eu-proposal-prevention-terrorist-content-online-important-mutation-e-commerce> accessed 10 March 2019

Barendt E, *Freedom of Speech* (OUP 2005)

Barrett PM and Sims JG, 'False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives' (*NYU Stern Center for Business and Human Rights,* February 2021)

Bates EP and others, *Law of the European Convention on Human Rights* (OUP 2009)

Beirich H, 'Resolute: The Battle Against Hate Demands Vigilance' (*Southern Poverty Law Center: Intelligence Report,* 10 September 2019) <https://www.splcenter.org/fighting-hate/intelligence-report/2019/resolute-battle-against-hate-demands-vigilance> accessed 20 November 2020

Ben-Zvi A, *The United States and Israel: The Limits of the Special Relationship* (Columbia University Press 1993)

Bennett WL and Livingston S, 'The disinformation order: Disruptive communication and the decline of democratic institutions' (2018) 33(2) EJ Comm 122

Benkler Y, Faris R and Roberts H, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (OUP 2018)

— — and others, 'Mail-In Voter Fraud: Anatomy of a Disinformation Campaign' (2020) Berkman Center Research Publication No 2020 <https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020> accessed 5 July 2021

Benvenisti E, 'Margin of Appreciation, Consensus, and Universal Standards, 31 NYUJ Intl L & Pol 843 (1999) 851-52

Berger JM, 'Nazis v. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks' (*George Washington University Center on Extremism,* September 2016) <https://cchs.gwu.edu/sites/g/files/zaxdzs2371/f/downloads/Nazis%20v.%20ISIS%20Final_0.pdf> accessed 13 May 2019.

Berlin I, *Four Essays on Liberty* (1st edn, Oxford Paperbacks 1969)

Black, Jr. CL, 'Forward: "State Action," Equal Protection, and California's Proposition 14' (1967) 81 Harv L Rev 69

Blank LR and Bayefsky AF (eds), *Incitement to Terrorism* (Brill Nijhoff 2018)

Bognar E, 'Digital News Report: Hungary' (*Reuters Institute,* 2020) <https://www.digitalnewsreport.org/survey/2020/hungary-2020/> accessed 7 June 2021

Brachotte G and others (eds), *Tweets from the Campaign Trail: Researching Candidates' Use of Twitter During the European Parliamentary Elections* (Peter Lang AG 2016)

Bradshaw S and Howard PN, 'Online Supplement to Working Paper 2018.1 Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation'

(*Oxford Internet Institute*, 2019) 29 <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct_appendix.pdf> accessed 6 June 2021
— — and others, 'Country Case Studies Industrialised Disinformation: 2020 Global Inventory of Organised Social Media Manipulation' (*Oxford Internet Institute*, 2020) <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/03/Case-Studies_FINAL.pdf> accessed 3 March 2021

Brauch JA, 'The Dangerous Search for and Elusive Consensus: What the Supreme Court Should Learn from the European Court of Human Rights' (2009) 52 Howard LJ 277

Brennan Center for Justice, 'It's Official: The Election Was Secure' (*Brennan Center for Justice,* 11 December 2020) <https://www.brennancenter.org/our-work/research-reports/its-official-election-was-secure> accessed 4 February 2021.
— — 'State Voting Bills Tracker: State lawmakers continue to introduce voting and elections bills at a furious pace' (*Brennan Center for Justice*, 28 May 2021) <https://www.brennancenter.org/our-work/research-reports/state-voting-bills-tracker-2021 > accessed 18 June 2021.

Bruhholm T and Johansen BS (eds), *Hate, Politics, Law: Critical Perspectives on Combating Hate* (OUP 2018)

Bures O, *EU Counterterrorism Policy* (Routledge 2011)

Busvine D, 'Google takes legal action over Germany's expanded hate-speech law,' *Reuters* (27 July 2021) <https://www.reuters.com/technology/google-takes-legal-action-over-germanys-expanded-hate-speech-law-2021-07-27/> accessed 21 July 2021.

Cannie H and Voorhoof D, 'The Abuse Clause and Freedom of Expression in the European Convention on Human Rights: An Added Value for Democracy and Human Rights Protection?' (2011) 29(1) NQHR 54

Casadevall J and others (eds), *Freedom of Expression: Essays in Honour of Nicolas Bratza* (Wolf Legal Publishers 2012)

Castor, Jr BL, Schoen D, and Van Deer Veen MT, 'Trial Memorandum of Donald J. Trump, 45th President of the United States of America' (Integral Text Publishing 2021)

Center for an Informed Public, Digital Forensic Lab, Graphika, and Stanford Observatory, 'The Long Fuse: Misinformation and the 2020 Election' (*Stanford Digital Repository*, 2021) <https://purl.stanford.edu/tr171zs0069> accessed 5 July 2021

Chemerinsky E, *Constitutional Law* (4th edn, Aspen Publishers 2013)
— — 'Rethinking State Action' (1985) 80 Northwest U L Rev 503

Chenoweth E and others (eds), *The Oxford Handbook on Terrorism* (OUP 2019)

Ciacchi AC, 'Egenberger and Comparative Law: A Victory of the Direct Horizontal Effect of Fundamental Rights' (2018) 5 EJCL 207

Cole D, 'The First Amendment's Borders: The Place of Holder v. Humanitarian Law Project in First Amendment Doctrine' (2016) 6 Harv L & Pol Rev 147

Common MF and Nielsen RK, 'Submission to UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on disinformation' (*Reuters Institute*, 15 February 2021) <https://reutersinstitute.politics.ox.ac.uk/risj-review/how-respond-disinformation-while-protecting-free-speech> accessed 5 July 2021.

Congressional Research Service, 'Report on US Foreign Aid to Israel' (RL33222, 16 November 2020) <https://fas.org/sgp/crs/mideast/RL33222.pdf > accessed 26 January 2020

Cox DA, 'After the ballots are counted: Conspiracies, political violence, and American exceptionalism', (*Survey Center on American Life,* 11 February 2021) <https://www.americansurveycenter.org/research/after-the-ballots-are-counted-conspiracies-political-violence-and-american-exceptionalism/> accessed 2 July 2021

Cram I (ed), *Extremism, Free Speech, and Counter-Terrorism Law and Policy* (Routledge 2019)

— — 'Coercing Communities or Promoting Civilised Discourse? Funeral Protests and Comparative Hate Speech Jurisprudence' (2012) 12 HRL Rev 455

Csaky Z, 'Nations in Transit 2020: Dropping the Democratic Façade in Europe and Eurasia' (*Freedom House,* 2020) 5 <https://freedomhouse.org/report/nations-transit/2020/dropping-democratic-facade> accessed 13 February 2021.

— — 'Nation in Transit 2021: The Antidemocratic Turn' (*Freedom House,* 2021) <https://freedomhouse.org/report/nations-transit/2021/antidemocratic-turn> accessed 12 May 2021.

Daly E, *Dignity Rights: Courts, Constitutions, and the Worth of the Human* Person (University of Pennsylvania Press 2013)

Deslatte M, 'Bill targeting critical race theory divides La. Lawmakers' *AP News* (27 April 2021) <https://apnews.com/article/education-race-and-ethnicity-government-and-politics-5c4a42900a3d17756d64b6d435f9f41f> accessed 29 June 2021.

Democracy Reporting International, 'Tackling Disinformation and Online Hate Speech: EU and Member State approaches, so far' (*Democracy Reporting International*, December 2020) <https://democracy-reporting.org/wp-content/uploads/2021/01/Tackling-Disinformation-and-Online-Hate-Speech-DRI.pdf> accessed 28 June 2021

Dworkin R, 'What is Equality? Part 3: The Place of Liberty' (1987) 73(1) Iowa L Rev 1

— — 'A New Map of Censorship' (*Index on Censorship*, May/June 1994)

— — *Sovereign Virtue: The Theory and Practice and Equality* (Harvard University Press 2000)

— — 'Even Bigots and Holocaust Deniers Must Have Their Say' *The Guardian* (14 February 2006) <http://www.theguardian.com/world/2006/feb/14/muhammadcartoons.comment> accessed 10 March 2019

Dzehtsiarou K, *European Consensus and the Legitimacy of the European Court of Human Rights* (CUP 2015)

— — and O'Mahony C, 'Evolutive Interpretation of Rights Provisions: A Comparison of the European Court of Human Rights and the U.S. Supreme Court' (2013) 44 Colum Hum Rts L Rev 309

EBU Research, 'Trust in Media 2020' (*EBU Research,* June 2020) <https://www.ebu.ch/publications/research/login_only/report/trust-in-media> accessed 25 June 2021

Edwards G and Meyer C, 'Introduction: Charting a Contested Transformation' (2008) 46(1) JCMS 1

Efrat A and others, 'Report on the Relationship Between Terrorist Threats and Governance Condition in the European Union' (*RECONNECT,* 29 June 2021) <https://reconnect-europe.eu/wp-content/uploads/2021/06/D11.3.pdf)> accessed 1 July 2020

Etinson A (ed), *Human Rights: Moral or Political* (OUP 2018)

EU Agency for Law Enforcement Cooperation, 'European Union Terrorism and Situation and Trend Report 2018' (*EU Agency for Law Enforcement Cooperation*, 2018) <https://op.europa.eu/en/publication-detail/-/publication/6a2c720a-93a3-11e8-8bc1-01aa75ed71a1/language-en> accessed 1 May 2019

Eurostat, 'First population estimates EU Population up to Almost 512 Million at 1 January 2017. Increase driven by migration' (*EuroStat Press Release*, 10 July 2017) <https://ec.europa.eu/eurostat/documents/2995521/8102195/3-10072017-AP-EN.pdf/a61ce1ca-1efd-41df-86a2-bb495daadbab> accessed 13 May 2019.

Fenwick H and Fenwick D, '"East"/"West" Divisions in Council of Europe States on Treatment of Sexual Minorities: The Response of Strasbourg Court and the Role of Consensus Analysis' (2019) 3 EHRLR 247

Fikfak V, 'Changing State Behaviour: Damages before the European Court of Human Rights' (2018) 29(4) EJIL 1091

Fisher C and Gaber I, '"Strategic Lying": The Case of Brexit and the 2019 U.S. Election' (2021) Int J of Press/Pol <doi:10.1177/1940161221994100>

Fiss J and Mchangama J, 'The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship' (*Justitia*, 2019) <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf> accessed 28 June 2021.

Føllesdal A, Peters B and Ulfstein G (eds), *Constituting Europe: The European Court of Human Rights in a National, European and Global Context* (CUP 2013)

Føllesdal A, 'Subsidiarity and International Human-Rights Courts: Respecting Self-Governance and Protecting Human Rights - or Neither?' (2016) 79 LCP 147

Fox News, 'Nunes Complaint - 3.18.19' <https://www.scribd.com/document/402297422/Nunes-Complaint-3-18-19> accessed 25 April 2019.

Frantziou E, *The Horizontal Effect of Fundamental Rights in the European Union: A Constitutional Analysis* (OUP 2019)

Fredman S, *Human Rights Transformed* (OUP 2008)

Freedom House, 'Freedom in the World 2020: United States (*Freedom House*, 2020) <https://freedomhouse.org/country/united-states/freedom-world/2020> accessed 2 July 2021

— — 'Freedom on the Net 2019: The Crisis of Social Media' (*Freedom House,* 2019) <https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf > accessed 10 February 2020.

Funk A and Shahbaz A, 'Freedom on the Net 2019: The Pandemic's Digital Shadow' (*Freedom House*, 2020) <https://freedomhouse.org/report/freedom-net/2020/pandemics-digital-shadow> accessed 25 March 2021.

Gardbaum S, 'The "Horizontal Effect" of Constitutional Rights' (2003) 102 Mich L Rev 387

Gearty C, 'The politics of terror' (*Index on Censorship Magazine,* 23 September 2015) <https://www.indexoncensorship.org/2015/09/the-politics-of-terror-conor-gearty/> accessed 20 June 2021).

Geiger AW, 'Key findings about the online news landscape in America' (*Pew Research Center,* 11 September 2019) <https://www.pewresearch.org/fact-tank/2019/09/11/key-findings-about-the-online-news-landscape-in-america/> accessed 16 June 2021.

George J, 'A Lesson in Critical Race Theory' (*American Bar Association*, 12 January 2021) <https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/civil-rights-reimagining-policing/a-lesson-on-critical-race-theory/> accessed 20 June 2021.

Gerards J, *General Principles of the European Convention on Human Rights* (CUP 2019)

— — 'Margin of Appreciation and Incrementalism in the Case Law of the European Court of Human Rights' (2018) 18 HRL Rev 495

Gilsinan K, 'Trump Keeps Invoking Terrorism to Get his Border Wall' (*The Atlantic,* 11 December 2019) <https://www.theatlantic.com/international/archive/2018/12/trump-incorrectly-links-immigration-terrorism/576358/> accessed 10 January 2020.

Gleicher N, 'Removing More Coordinated Inauthentic Behavior From Iran and Russia (*Facebook News,* 21 October 2019) < https://about.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-iran-and-russia/> accessed 25 January 2019.

Greene AR and Simpson RM, 'Tolerating Hate in the Name of Democracy' (2017) 80(4) MLR 746

Greer S, Gerards J and Slowe R, *Human Rights in the Council of Europe and the European Union* (CUP 2018).
— — *The Margin of Appreciation: Interpretation and Discretion under the European Convention on Human Rights* (Council of Europe Publishing 2000)

Hare I and Weinstein J (eds), *Extreme Speech and Democracy* (OUP 2009)

Harris DJ and others, *Harris, O'Boyle, and Warbrick's Law of the European Convention on Human Rights* (4th edn, OUP 2018)

Hauss B, 'Laws Suppressing Boycotts of Israel Don't Prevent Discrimination – They Violate Civil Liberties' (*ACLU*, 22 February 2019) <https://www.aclu.org/blog/free-speech/laws-suppressing-boycotts-israel-dont-prevent-discrimination-they-violate-civil> accessed 28 June 2021

Heinze E, *Hate Speech and Democratic Citizenship* (OUP 2016)

Heller B and others, 'Freedom of Expression: A Comparative Summary of United States and European Law' (*Transatlantic Working Group*, 3 May 2019)

Herz M and Molnar P (eds)*, The Content and Context of Hate Speech: Rethinking Regulation and Responses* (CUP 2012)

Heyman SJ, 'Positive and Negative Liberty' (1992) 68 Chi-Kent L Rev 8
— — *Free Speech and Human Dignity* (Yale University Press 2008)

Hindle G and others, 'New Global Data on Political Parties: V-Party (Briefing Paper No 9)' (*V-Dem Institute,* 26 October 2020) <https://www.v-dem.net/media/filer_public/b6/55/b6553f85-5c5d-45ec-be63-a48a2abe3f62/briefing_paper_9> accessed 3 January 2021

Hmielowski J, Hutchens M and Beam M, 'Conservatives are more likely than liberals to exist in a media echo chamber' (*London School of Economics Blog,* 22 December 2020) <https://blogs.lse.ac.uk/usappblog/2020/12/22/conservatives-are-more-likely-than-liberals-to-exist-in-a-media-echo-chamber/> accessed 3 February 2021

van Hoboken J and Keller D, 'Design Principles for Intermediary Liability Laws' (8 October 2019) Transatlantic Working Papers Series, 3 - 4 <https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/Intermediary_Liability_TWG_van_Hoboken_Oct_2019.pdf> accessed 10 July 2021
— — 'The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications' (*Transatlantic Working Group*, 3 May 2019) <https://www.ivir.nl/publicaties/download/TERREG_FoE-ANALYSIS.pdf> accessed 11 March 2020

Holmes OW, 'Law in Science and Science in Law' (1899) 12(7) Harv L Rev 443

Hudson Jr. DL, 'Free speech or censorship? Social media litigation is a hot legal battleground' (*ABA Journal* 1 April 2019) <https://www.abajournal.com/magazine/article/social-clashes-digital-free-speech> accessed 11 July 2021.

Human Rights Watch, 'Germany: Flawed Social Media Law–NetzDG is Wrong Response to Online Abuse' (*Human Rights Watch,* 14 February 2018) <https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law> accessed 28 June 2021

— — 'US: States Use Anti-Boycott Laws to Punish Responsible Businesses' (*Human Rights Watch*, 23 April 2019) <https://www.hrw.org/news/2019/04/23/us-states-use-anti-boycott-laws-punish-responsible-businesses> accessed 30 November 2020

Ipsos, 'Ipsos/Reuters Poll: The Big Lie – Over half of Republicans believe Donald Trump is the actual President of the United States' <https://www.ipsos.com/sites/default/files/ct/news/documents/2021-05/Ipsos%20Reuters%20Topline%20Write%20up-%20The%20Big%20Lie%20-%2017%20May%20thru%2019%20May%202021.pdf> accessed 2 July 2021

Izaguirre A, 'GOP pushes bills to allows social media 'censorship' lawsuits' *AP News* (7 March 2021) <https://apnews.com/article/donald-trump-legislature-media-lawsuits-social-media-848c0189ff498377fbfde3f6f5678397> accessed 10 March 2021.

Jackson C and Silverstein K, 'Over half of Republicans believe Donald Trump is the actual President of the United States' (*IPSOS,* 21 May 2021) <https://www.ipsos.com/en-us/news_and_polls/over-half-republicans-believe-donald-trump-actual-president-united-states> accessed 8 July 2021.

Jacobs L and Van Spanje J, 'Prosecuted, yet popular? Hate Speech Prosecution of anti-immigration politicians in the news and electoral support' (2020) 18 Comp Eur Polit 899

Jeffries I, 'Russia's Constitutional Amendment from an International Law Perspective' (*Public International Law & Policy Group,* 1 March 2021)

Jigsaw Research, 'News Consumption in the UK: 2019' (*Ofcom,* 24 July 2019) <https://www.ofcom.org.uk/__data/assets/pdf_file/0027/157914/uk-news-consumption-2019-report.pdf> accessed 2 June 2020.

Jones K, 'Online Disinformation and Political Discourse: Applying a Human Rights Framework' (*Chatham House International Law Program*, 6 November 2019) 2 <https://www.chathamhouse.org/2019/11/online-disinformation-and-political-discourse-applying-human-rights-framework> accessed 3 July 2021.

Jurkowitz M and others, 'U.S. Media Polarization and the 2020 Election: A Nation Divided' (*Pew Research Center,* 24 January 2020) <https://www.journalism.org/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/> accessed 8 June 2020.

Kaplan T, 'How the Trump Campaign Used Facebook Ads to Amplify His 'Invasion' Claim' *The New York Times* (5 August 2019) <https://www.nytimes.com/2019/08/05/us/politics/trump-campaign-facebook-ads-invasion.html> accessed 12 December 2019.

Keller D, ‘Internet Platforms: Observations on Speech, Danger, and Money’ (2018) Hoover Institution’s Aegis Paper Series No 1807 <https://ssrn.com/abstract=3262936> accessed 28 June 2021
— — ‘Internet Platforms: Observations on Speech, Danger, and Money,’ Hoover Institution’s Aegis Paper Series, No. 1807 (13 June 2018)
— — ‘Who Do You Sue? State And Platform Hybrid Power Over Online Speech’, A Hoover Institution Essay, Aegis Series Paper No. 1902 (29 January 2019)
— — ‘The EU’s Terrorist Content Regulation: Expanding The Rule of Platform Terms of Service and Exporting Expression Restrictions From the EU’s Most Conservative Member States’ (*Center for Internet and Society,* 25 March 2019) <http://cyberlaw.stanford.edu/blog/2019/03/eus-terrorist-content-regulation-expanding-rule-platform-terms-service-and-exporting> accessed 20 January 2019.
— — and Joan Barata, ‘Regulating Online Terrorist Content: A Discussion with Stanford CIS experts about New EU Proposals’ (*Stanford Law School*, 25 April 2019) <https://law.stanford.edu/2019/04/25/regulating-online-terrorist-content-a-discussion-with-stanford-cis-directors-about-new-eu-proposals/> accessed 4 November 2019.
— — ‘Amplification and Its Discontents’, (*Knight First Amendment Institute*, 8 June 2021) <https://knightcolumbia.org/content/amplification-and-its-discontents> accessed 6 July 2021

Keszthelyi C, ‘Xenophobia Skyrocketing in Hungary, Surveys Reveal’ (*Budapest Business Journal,* 17 November 2016) <https://bbj.hu/budapest/xenophobia-skyrocketing-in-hungary-surveys-reveal_124920> accessed 3 November 2020

Kiska R, 'Hate Speech: A Comparison between the European Court of Human Rights and the United States Supreme Court Jurisprudence' (2012) 25 Regent U L Rev 107

Kuhn PY, ‘Reforming the Approach to Racial and Religious Hate Speech under Article 10 of the European Convention on Human Rights’ (2019) 19 HRL Rev 119

Kamber K, *Prosecuting Human rights Offences: Rethinking the Sword Function of Human Rights Law* (BRILL 2017)

Kant I, *Groundwork of the Metaphysics of Morals* (1979) (Mary Gregor tr, 1st edn, 1998)

Kaye D, Cannataci J and Aoláin FN, ‘Letter to the European Union’ (OL OTH 71/2018, 7 December 2018) <https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24234> accessed 9 April 2019
— — ‘Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (14 February 2019) OL USA 2/2019 <https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24338> accessed 30 November 2020

Lavrysen L and Mavronicola N (eds), *Coercive Human Rights: Positive Duties to Mobilise the Criminal Law under the ECHR* (Hart Publishing 2020).

Leczykiewicz D, ‘The Judgment in Bauer and the Effect of the EU Charter of Fundamental Rights in Horizontal Situations’ (2020) 16 ERCL 323

Lee D, 'Germany's NetzDG and the Threat to Online Free Speech (*Media Freedom & Information Access Clinic at Yale Law School*, 10 October 2017) <https://law.yale.edu/MFIA/CASE-DISCLOSED/GERMANYS-NETZDG-AND-THREAT-ONLINE-FREE-SPEECH> accessed 11 March 2020

Leersson P and Tworek H, 'An Analysis of Germany's NetzDG Law' (*Transatlantic Working Group*, 15 April 2019) 1 <https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf> accessed 28 June 2021.

Legg A, *The Margin of Appreciation in International Human Rights Law Deference and Proportionality* (OUP 2012)

Letsas G, *A Theory of Interpretation of the European Convention on Human Rights* (OUP 2009)
— — 'Two Concepts of the Margin of Appreciation' (2006) 26(4) OJLS 705

Levin S, ''It's a small group of people': Trump again denies white nationalism is a rising threat' *The Guardian* (15 March 2019) <https://www.theguardian.com/us-news/2019/mar/15/donald-trump-denies-white-nationalism-threat-new-zealand> accessed 14 April 2019

Linsday JM, 'The 2020 Election by the Numbers' (*Council on Foreign Relations,* 15 December 2020) <https://www.cfr.org/blog/2020-election-numbers> accessed 14 February 2020.

Little D, 'The Making of a Special Relationship: The United States and Israel, 1957-68' (1993) 25 Int J Middle East Stud 563

Locke J, *Second Treatise of Government* (1689) (CB Macpherson ed, 1st edn 1980)

de Londras F and Dzehtsiarou K, 'Mission Impossible? Non-Execution Through Infringement Proceedings in the European Court of Human Rights (2017) 66 ICLQ 467

Lopez S and Tager J, 'Truth on the Ballot: Fraudulent News, the Midterm Elections, and Prospects for 2020' (*Pen America,* 13 March 2019) <https://pen.org/wp-content/uploads/2019/03/Truth-on-the-Ballot-report.pdf> accessed 20 June 2021.

Lyons K, 'Christchurch mosque attacks: suspect charged with "terrorist act"' *The Guardian* (21 May 2019) accessed 20 January 2020

Machan TR (ed), *Liberty and Equality* (Hoover Institution Press 2002)

Maitra I and McGowan MK, 'On Racist Hate Speech and the Scope of a Free Speech Principle' (2007) 23(2) CJLJ 343

Mälksoo L, 'Russian's Constitutional Court Defies the European Court of Human Rights' (2016) 12(2) Eur Const Law Rev 377

Mallow T and Schwartz D, '74% of voters say democracy in the US is under threat, Quinnipiac University national poll finds; 52% say president trump should be

removed from office' (*Quinnipiac University*, 11 January 2021) <https://poll.qu.edu/images/polling/us/us01112021_usmk38.pdf>.

Masterman R and Schütze R (eds), *The Cambridge Companion to Comparative Constitutional Law* (CUP 2019)

Matsa KE, 'Across Western Europe, public news media are widely used and trusted sources of news' (*Pew Research Center,* 8 June 2018) <https://www.pewresearch.org/fact-tank/2018/06/08/western-europe-public-news-media-widely-used-and-trusted/> accessed 15 March 2021

Marchel N and others, 'Polarization, Partisanship and Junk News Consumption on Social Media During the 2018 US Midterm Elections' (*Oxford Internet Institute,* 1 November 2018) <https://demtech.oii.ox.ac.uk/research/posts/polarization-partisanship-and-junk-news-consumption-on-social-media-during-the-2018-us-midterm-elections/> accessed 28 January 2021.

Matsuda M and others (eds), *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Westview Press 1993)

McCabe D, 'Florida, in a First, Will Fine Social Media Companies That Bar Candidates' *The New York Times* (24 May 2021) <https://www.nytimes.com/2021/05/24/technology/florida-twitter-facebook-ban-politicians.html> accessed 1 July 2021.

McCrudden C, 'Human Dignity and Judicial Interpretation of Human Rights' (2018) 19 EJIL 4

Mcdonald RSJ, Matscher F and Petzold H (eds), *The European System for the Protection of Human Rights* (Martinus Nijhoff 1993)

McGonagle T, 'The Council of Europe against online hate speech: Conundrums and challenges (Expert Paper)' <https://rm.coe.int/16800c170f> accessed 17 June 2021

Meiklejohn A, *Free Speech and its Relation to Self-Government* (Harper Brothers Publishers 1948)
— — *Political Freedom: The Constitutional Powers of the People* (Harper Publishing 1960)

Minow M, 'Alternatives to the State Action Doctrine in the Era of Privatisation, Mandatory Arbitration, and the Internet: Directing Law to Serve Human Needs,' (2017) 52 Harv Civil Rights-Civil Liberties L Rev 145

Mitchell A others, 'Americans Who Mainly Get Their News on Social Media are Less Engaged, Less Knowledgeable' (*Pew Research Center,* 30 July 2020) <https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/> accessed 30 June 2021

Möller K, *The Global Model of Constitutional Rights* (OUP 2015)

Molnar P and Mchangama, 'The Problem with Hate Speech Laws' (2015) 13(1) Rev Faith and Int Affrs 75

Monmouth University, ‘Majority Back Capitol Riot Commission’ (*Monmouth University Polling Institute,* 17 March 2021) <https://www.monmouth.edu/polling-institute/reports/MonmouthPoll_US_031721/> accessed 8 July 2021.

Moon R, *Putting Faith in Hate* (CUP 2018)

Moore M and Tambini D (eds), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (OUP 2018)

de Morree P, *Rights and Wrongs under the ECHR: The prohibition of abuse of rights in Article 17 of the European Convention on Human Rights* (Intersentia 2016)

Mowbray A, ‘Subsidiarity and the European Convention on Human Rights’ (2015) 15 HRL Rev 313

Mueller J and Stewart MG, *Terror, Security, and Money: Balancing the Risks, Benefits, and Costs of Homeland Security* (OUP 2011)

— — ‘Public Opinion and Counterterrorism Policy (White Paper)' (*The Cato Institute,* 20 February 2018) <https://www.cato.org/publications/white-paper/public-opinion-counterterrorism-policy> accessed 5 May 2019

Murphy CC, ‘Using the EU Charter of Fundamental Rights Against Private Parties after Association De Médiation Sociale’ (2014) 2 EHRLR 170

Narveson J and Sterba JP (eds), *Are Liberty and Equality Compatible* (CUP 2010)

Nelson A, ‘European parties urged to agree Israel boycott tactics are antisemitic’ (*The Guardian,* 24 October 2018) <https://www.theguardian.com/world/2018/oct/24/european-parties-urged-agree-israel-boycott-bds-antisemitic-mep> accessed 14 April 2019.

New York Times Magazine*,* ‘The 1619 Project’ <https://www.nytimes.com/interactive/2019/08/14/magazine/1619-america-slavery.html> accessed 29 June 2021.

Nolan A, ‘Holding non-state actors to account for constitutional economic and social rights violations: Experience and lessons from South Africa and Ireland’ (2014) 12(1) ICON 61

Oxford University Press ‘Word of the Year 2016’ <https://languages.oup.com/word-of-the-year/2016/> accessed 20 May 2021

Phillipson G and Williams A, ‘Horizontal effect and the constitutional constraint’ (2011) 74(6) MLR 878

Pickard V, ‘The Strange Life and Death of the Fairness Doctrine: Tracing the Decline of Positive Freedoms in American Policy Discourse’ (2018) 12 Int J Com 3434

Pinto M, ‘Historical Trends of Human Rights Gone Criminal (2020) 42(:4) HRQ 729

Pollicino O and Bietti E, 'Truth and Deception across the Atlantic: A Roadmap of Disinformation in the US and Europe' (2019) 11 Italian J Pub L 43

Post R, 'Participatory Democracy and Free Speech' (2011) 97(3) Va L Rev 477
— — 'Legitimacy and Hate Speech' (2017) 32 Const Comment 651

Quinnipiac University, Data for Progress and Vox, 'Poll on Election Trust' (*Quinnipiac University*, 2021) <https://www.filesforprogress.org/datasets/2021/1/dfp_vox_election_trust.pdf>).

Rashawn R, 'Addressing hate crimes in Maryland and in America,' (*Brookings Institution*, 7 April 2020) <https://www.brookings.edu/testimonies/addressing-hate-crimes-in-maryland-and-in-america/> accessed 12 November 2020

Reimann M and Zimmeran R (eds), *The Oxford Handbook on Comparative Law* (OUP 2019)

Repucci S, 'Media Freedom: A Downward Spiral' (*Freedom House*, 2019) <https://freedomhouse.org/report/freedom-and-media/2019/media-freedom-downward-spiral> accessed 8 June 2021.

Ripstein A (ed), *Ronald Dworkin* (CUP 2007)

Rosenfeld M (ed), *The Oxford Handbook of Comparative Constitutional Law* (OUP 2012)
— — 'Comparing constitutional review by the European Court of Justice and the U.S. Supreme Court' (2006) 4(4) ICON 618

Rudd A, 'Speech at Digital Forum' (San Francisco, 13 February 2018)

Rumero L, 'Terrorism Is Surging in the U.S., Fueled by Right-Wing Extremist Ideologies' (*Quartz,* 17 August 2018) <https://qz.com/1355874/terrorism-is-surging-in-the-us-fueled-by-right-wing-extremists/> accessed 25 April 2019.

Ryan A (ed), *The Idea of Freedom: Essays in Honour of Isaiah Berlin* (OUP 1979)

Saul B, *Defining Terrorism in International Law* (OUP 2010)
— — 'Defining Terrorism: A Conceptual Minefield' (2015) Sydney Law School Legal Studies Research Paper No 15/84 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2664402> accessed 13 May 2019

Schauer F, 'Is it Better to Be Safe than Sorry? Free Speech and the Precautionary Principle' (2009) 36(2) Pepperdine L Rev 301
— — *Free Speech: a philosophical enquiry* (CUP 1982)
— — 'Free Speech and Obedience to Law' (2017) 32 Const Comment 661

Smith RC, 'Fake news, French law and democracy: lessons for the United Kingdom?' (2019) 11(1) J of Med L 52

Sottiaux S, 'Leroy v France: Apology of Terrorism and the Malaise of the European Court of Human Rights' Free Speech Jurisprudence' (2009) EHRLR 415

Spielman D, 'Allowing the Right Margin in the European Court of Human Rights and the National Margin of Appreciation Doctrine: Waiver or Subsidiarity of European Review?' (2012) 14 CYELS 381, updated at <http://www.echr.coe.int/Documents/Speech_20140113_Heidelberg_ENG.pdf>

Statista, 'The world leaders with the most Twitter followers 2020' (*Statista*, 27 January 2021) <https://www.statista.com/statistics/281375/heads-of-state-with-the-most-twitter-followers/> accessed 2 May 2021

Stevens E, *US policy towards Israel: the role of political culture in defining the 'special relationship'* (Sussex Academic Press 2006)

Stone A and Schauer F (eds), *The Oxford Handbook on Freedom of Speech* (OUP 2020)

Strossen N, *HATE: Why We Should Resist It with Free Speech, Not Censorship* (OUP 2018)

Sykes E and Hinger S, 'State Lawmakers are Trying to Bank Talk About Race in Schools' (*ACLU,* 14 May 2021) <https://www.aclu.org/news/free-speech/state-lawmakers-are-trying-to-ban-talk-about-race-in-schools/> accessed 29 June 2021.

Tankovska H, 'The world leaders with the most Twitter followers 2020' (*Statista*, 27 January 2021) <https://www.statista.com/statistics/281375/heads-of-state-with-the-most-twitter-followers/> accessed 3 February 2021

Tulkens F, 'The Paradoxical Relationship between Criminal Law and Human Rights (2011) 9 J Intl Crim Just 577

Tulkens F, 'When to say is to do: freedom of expression and hate speech in the case-law of the European Court of Human Rights', ECtHR European Judicial Training Network: Seminar on Human Rights for European Judicial Trainers (Strasbourg, 9 October 2012)

Tushnet M, 'The Issue of State Action/Horizontal Effect in Comparative Constitutional Law' (2003) 1 ICON 79

Ulfstein G, 'The European Court of Human Rights as a Constitutional Court?' (2014) PluriCourts Research Paper No. 14-08

University of Massachusetts, 'University of Massachusetts Amherst/WCVB National Poll of President Biden's First 100 Days' <https://polsci.umass.edu/sites/default/files/Biden100DaysToplines%20%281%29.pdf> accessed 2 July 2021

University of Oxford, 'Comparative Hate Speech Law: Annexure (*Oxford Pro Bono Publico*, March 2012) <https://www.law.ox.ac.uk/sites/files/oxlaw/1a._comparative_hate_speech_annex.pdf> accessed 1 September 2020

Voorhoof D, 'The European Convention on Human Rights: The Right to Freedom of Expression and Information Restricted by Duties and Responsibilities in a Democratic Society' (2015) 7 JOUR 1

Waldron J, 'The Oliver Wendell Holmes Lectures, Dignity and Defamation: The Visibility of Hate' (2009) 123 Harv L Rev 1596

—— *The Harm in Hate Speech* (Harvard University Press 2012)

van der Walt J, *The Horizontal Effect Revolution and the Question of Sovereignty* (De Gruyter 2014)

Wardle C, 'Fake News. It's Complicated' (*First Draft,* 16 February 2018) <https://firstdraftnews.org/articles/fake-news-complicated/> accessed 21 May 2021.

Weinstein J, *Hate Speech, Pornography, and Radical Attacks on Free Speech Doctrine* (Routledge 1999)
— — 'Participatory Democracy as the Central Value of American Free Speech Doctrine' (2011) 97 Va L Rev 491
— — 'Hate Speech Bans, Democracy, and Political Legitimacy' (2017) 32 Const Comment 527
— — 'Viewpoint Discrimination, Hate Speech, and Political Legitimacy: A Reply' (2017) 32 Const Comment 715

West DM, 'How to combat fake news and disinformation' (*Brooking Institution*, 18 December 2017) <https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/> accessed 13 June 2021

Wilkinson College, 'Survey of American Fears 2017: Fear of Extremism and the Threat to National Security' (*Chapman University: The Voice of Wilkinson,* 11 October 2017) <https://blogs.chapman.edu/wilkinson/2017/10/11/fear-of-extremism-and-the-threat-to-national-security/> accessed 5 May 2019

Witz B, 'This Time, Colin Kaepernick Takes a Stand by Kneeling,' *The New York Times* (1 September, 2016) <https://www.nytimes.com/2016/09/02/sports/football/colin-kaepernick-kneels-national-anthem-protest.html> accessed 12 July 2020

Wu T, 'Is the First Amendment Obsolete?' (2018) 117(3) Mich L Rev 547

YouGov, '*The Economist*/YouGov Poll, 22-25 May 2021' <https://docs.cdn.yougov.com/pxuc7wjg52/econTabReport.pdf> accessed 2 July 2021
— — Fourteen percent of total respondents and only seven percent of Democrats believed the attack was justified (YouGov, 'Yahoo! News COVID-19 Vaccination Survey - 20210526 (24-26 May 2021)' <https://docs.cdn.yougov.com/zjdg6ujrzh/20210526_yahoo_vaccine_tabs.pdf> accessed 2 July 2021

Young EA, 'What Can Europe Tell Us About the Future of American Federalism?' (2017) 49 Ariz St L J 1109

Zedner L and Roberts JV (eds), *Principals and Values in Criminal Law and Criminal Justice* (OUP 2012)

Zeldin L, 'Rep. Zeldin Leads 60 Members in Introducing Israel Anti-Boycott Act (Press Release)' (14 January 2020) <https://zeldin.house.gov/media-center/press-releases/rep-zeldin-leads-60-members-introducing-israel-anti-boycott-act> accessed 29 June 2021

Zysset A, 'Searching for the Legitimacy of the European Court of Human Rights: The Neglected Role of "Democratic Society"' (2016) 5(1) GlobCon 16