View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Helsingin yliopiston digitaalinen arkisto

Cooperation, Contest, and Co-optation:

Freedom of Expression and Positive Obligations in EU Social Media Content Regulation

Miikka Hiltunen
014456624

Pro gradu -tutkielma
Helsingin yliopisto
Oikeustieteellinen tiedekunta
Valtiosääntöoikeus; OTM-836S
Ohjaaja: Tuomas Ojanen
4 elokuuta 2020

Tiedekunta – Fakultet – Faculty Oikeustieteellinen tiedekunta		Koulutusohjelma – Utbildingsprogram – Degree Programme Oikeustieteen maisterin koulutusohjelma
Tekijä – Författare – Author Miikka Hiltunen		
Työn nimi – Arbetets titel – Title Cooperation, Contest, and Co-optation: Freedom of Expression and Positive Obligations in EU Social Media Content Regulation		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track valtiosääntöoikeus		
Työn laji – Arbetets art – Level Pro gradu -tutkielma	Aika – Datum – Month and year 4.8.2020	Sivumäärä – Sidoantal – Number of pages 82

Tiivistelmä – Referat – Abstract

The thesis analyzes how the EU regulates freedom of expression on social media and how the regulation transfers power. It interrogates EU's positive fundamental right obligation to put in place a legislative and administrative framework to prevent and redress different online harms. By taking up a critical method that analyzes the contradiction between efficiency and democracy, the thesis focuses on the changes that are happening to positive obligations as part of the fundamental rights structure. The thesis links the structural changes to the advancement of neoliberal governmentality that favors managerial techniques such as cost-benefit analysis and privatization in government.

After laying out the foundations, the thesis is divided into two parts. The first part departs from the observation that the EU has pressure to shift positive obligations regarding the protection of people's freedom of expression and related rights to social media companies. It develops the general analytical framework of cooperation and contest which is informed by power struggles. In cooperation, the interests of the EU and the companies in rights protection are considered aligned. In contest, the focus is on situations where the interests diverge and where the EU and the companies deploy their respective strategies to assert their power unilaterally. The second part contextualizes the framework of cooperation and contest in three case studies. The case studies analyze three different initiatives to regulate social media introduced by the EU between 2016 and 2019. The initiatives are EU Code of conduct on hate speech, the revised Directive on audiovisual media services, and the Regulation on preventing the dissemination of terrorist content online. The case studies locate cooperation and contest in regulation through the analysis of relevant policy documents and other preparatory materials of the legislative processes.

The thesis concludes that the emerging regulatory framework for rights protection exhibits hybridity that results from the interconnectedness of public and private power. While the framework can provide more effectiveness for the EU in protecting rights and public interest on social media, by deploying managerial techniques it also tends to sideline the considerations for people's democratic self-determination, and reinforce the power of the executive and large social media companies. It is argued that new ways to enhance the horizontality of rights are needed for people to assert their rights against emerging hybrid power.

Avainsanat – Nyckelord – Keywords
Freedom of expression, Positive obligations, Neoliberalism, Social media, Platform power

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Tuomas Ojanen

Säilytyspaikka – Förvaringställe – Where deposited

Helsingin yliopiston kirjasto, Helsingfors universitets bibliotek, Helsinki University Library

Muita tietoja – Övriga uppgifter – Additional information

Contents

Bibliography

Primary Sources

EU legislation

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

Code of conduct on countering illegal hate speech online (2016) <https://bit.ly/3i3Jzzh> accessed 22 May 2020

Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C326/47

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 and national laws transposing it [2008] OJ L328/55

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') [2000] OJ L178/1

Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities [2018] OJ L303/69

EU Code of Practice on Disinformation (2018) <https://bit.ly/33hRTHt> accessed 19 July 2020

Interinstitutional agreement on better law-making [2003] OJ C321/1

Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European Commission on Better Law-Making [2016] OJ L123/1

Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA [2016] OJ L135/53

ECHR

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

Case-law

ECtHR

Case "relating to certain aspects of the laws on the use of languages in education in Belgium" v Belgium (Belgian Linguistic case) App nos 1474/62 to 2126/64 (ECtHR, 23 July 1968)

Handyside v the United Kingdom App no 5493/72 (ECtHR, 7 December 1976)

Airey v Ireland App no 6289/73 (ECtHR, 9 October 1979)

Marckx v Belgium App no 6833/74 (ECtHR, 13 June 1979)

Young, James and Webster v The United Kingdom App nos 7601/76 and 7806/77 (ECtHR, 13 August 1981)

X and Y v the Netherlands App no 8978/80 (ECtHR, 26 March 1985)

Rees v the United Kingdom App no 9532/81 (ECtHR, 17 October 1986)

Plattform "Ärtze für das Leben" v Austria App no 10126/82 (ECtHR, 21 June 1988)

Costello-Roberts v the United Kingdom App no 13134/87 (ECtHR, 25 March 1993)

Jersild v Denmark App no 15890/89 (ECtHR, 23 September 1994)

Osman v the United Kingdom ECHR 1998-VIII 101

Özgür Gündem v Turkey ECHR 2000–III 1

VgT Verein Gegen Tierfabriken v Switzerland ECHR 2001–VI 243

Appleby and Others v the United Kingdom ECHR 2003–VI 165

Assanidze v Georgia ECHR 2004–II 221

Öneryildiz v Turkey ECHR 2004–XII 79

Sychev v Ukraine App no 4773/02 (ECtHR, 11 October 2005)

Storck v Germany ECHR 2005–VI 111

Anheuser-Busch Inc v Portugal ECHR 2007–I 39

CJEU

Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV*, EU:C:2012:85

Case C-286/12 *European Commission v Hungary*, EU:C:2012:687

Case C-617/10, *Åklagaren v Hans Åkerberg Fransson*, EU:C:2013:105

Case C-270/12 *United Kingdom of Great Britain and Northern Ireland v European Parliament and Council of the European Union (ESMA)*, EU:C:2014:18

Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others*, EU:C:2014:238

Case C-131/12 *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, EU:C:2014:317

Case C-362/14 *Maximillian Schrems v Data Protection Commissioner*, EU:C:2015:650

Case C-205/15 *Direcția Generală Regională a Finanțelor Publice Brașov (DGRFP) v Vasile Toma and Biroul Executorului Judecătoresc Horațiu-Vasile Cruduleci*, EU:C:2016:499

Case C-619/18 *European Commission v Poland*, EU:C:2019:531

Case C-507/17 *Google LLC, successor in law to Google Inc v Commission nationale de l'informatique et des libertés (CNIL)*, EU:C:2019:772

Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited*, EU:C:2019:821

Case C-311/18 *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems*, EU:C:2020:559

EU policy documents

European Commission

Commissioner for Justice, Consumers and Gender Equality, 'Countering illegal hate speech online – EU Code of Conduct ensures swift response' IP/19/805

Commissioner for Justice, Consumers and Gender Equality and Directorate-General for Justice and Consumers, 'Code of Conduct on countering illegal hate speech online: Fourth evaluation confirms self-regulation works' (Factsheet 4th round) (February 2019) <https://bit.ly/3gmNlTD> accessed 26 July 2020

Commissioner for Justice, Consumers and Gender Equality and Directorate-General for Justice and Consumers, 'How the Code of Conduct helped countering illegal hate speech online' (February 2019) <https://bit.ly/33fgsVv> accessed 26 July 2020

European Commission, 'Behavioural Study on Advertising and Marketing Practices in Online Social Media' (Final Report, June 2018)

European Commission, 'Better Regulation Toolbox: TOOL #18 – The choice of policy instruments' <https://ec.europa.eu/info/files/better-regulation-toolbox-18_en> accessed 18 June 2020

European Commission, 'Better Regulation Toolbox, Tool #57: Analytical methods to compare options or assess performance' <https://bit.ly/2XjYVaO> accessed 25 July 2020

European Commission, 'Code of Conduct on countering illegal hate speech online: Questions and answers on the fourth evaluation' MEMO/19/806

European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Preventing radicalisation to terrorism and violent extremism: strengthening the EU's response' COM (2013) 941 final

European Commission, 'Digital Single Market: updated audiovisual rules' MEMO/18/4093

European Commission, 'Follow up to the European Parliament legislative resolution on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' SP (2019) 440

European Commission, 'Impact assessment accompanying the document Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities' SWD (2016) 168 final

European Commission, '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' SWD (2018) 408 final

European Commission, 'Launch of the EU Internet Forum Brussels, 3 December 2015' (Redacted), Ref Ares(2016)1160160 - 07/03/2016 <https://bit.ly/314bsjy> accessed 22 May 2020

European Commission, 'Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities' COM (2016) 287 final

European Commission, ‘Proposal of a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online’ COM (2018) 640 final

European Commission, ‘Revision of the Audiovisual Media Services Directive (AVMSD)’ (5 December 2018) <https://bit.ly/3194FoR> accessed 6 June 2020

European Commission, ‘Shaping the Digital Single Market’ (last update 19 February 2020) <https://bit.ly/3gl6mWQ> accessed 2 June 2020.

European Commission DG Home Affairs, ‘Jour Fixe Meeting: EU Internet Forum’ (Redacted) (10 June 2015) 3 <https://bit.ly/2Db1zZC> accessed 22 May 2020

European Commission Directorate-General Migration and Home Affairs, ‘Launch of EU Internet Forum: Summary of Meeting’ (Redacted) Ref Ares(2016)1158809 (7 March 2016) <https://bit.ly/3gnrQlU> accessed 22 May 2020

European Commission Press Release, ‘EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online’ IP/15/6243.

European Commission Press Release, ‘EU Internet Forum: A major step forward in curbing terrorist content on the internet’ IP/16/4328

European Commission Press Release, ‘Fighting Terrorism Online: Internet Forum pushes for automatic detection of terrorist propaganda’ IP/17/5105

European Commission Secretary General, Decision of the Secretary General on behalf of the Commission pursuant to Article 4 of the implementing rules to Regulation (EC) N° 1049/20011, C (2016) 2065 final <https://bit.ly/3fgbJFm> accessed 22 May 2020

Council of the European Union

Council of the European Union, ‘Assessment of the Code of Conduct on Hate Speech online: State of Play’ (27 September 2019) 12522/19

Council of the European Union, ‘Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities (First reading) – general approach’ (24 May 2017) 9691/17

Council of the European Union, ‘Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online – general approach’ (3 December 2018) 14978/18

EU Counter-Terrorism Coordinator, ‘EU CTC input for the preparation of the informal meeting of Justice and Home Affairs Ministers in Riga on 29 January 2015’ (17 January 2015) DS 1035/15

European Parliament

European Parliament, ‘Draft Report with recommendations to the Commission on Digital Services Act: Improving the functioning of the Single Market’ 2020/2018 (INL)

European Parliament legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, P8_TA (2019) 0421

European Parliament, ‘Report on competition policy – annual report 2019’ P9_TA (2020) 0158

European Parliamentary Research Service, ‘The Audiovisual Media Services Directive: State of Play’ (November 2015) PE 571.329

European Parliamentary Research Service, ‘Protecting the Rule of Law in the EU: Existing Mechanisms and Possible Improvements’ (November 2019) PE 642.280

Other EU documents

EDiMA, ‘Consultation on Directive 2010/13/EU on audio-visual media services (AVMSD) A media framework for the 21st century: Questionnaire’ (July–September 2015) 17 <https://bit.ly/33kKzuK> accessed 8 June 2020

ERGA (European Regulators Group for Audiovisual Media Services), ‘Analysis & Discussion Paper to contribute to the consistent implementation of the revised Audiovisual Media Services (AVMS) Directive’ (2018) <https://erga-online.eu/?page_id=14> accessed 25 July 2020

EU Internet Referral Unit, ‘Transparency Report 2018’ (20 December 2019) <https://bit.ly/33i15LS> accessed 15 May 2020

European Ombudsman, Decision in case 292/2016/AMF on the European Commission´s handling of a request for access to documents related to the launch of the EU Internet Forum (5 July 2017)

Fiedler K (EDRi), ‘Complaint about maladministration’ (17 February 2016) <https://bit.ly/2Dohzaq> accessed 22 May 2020

FRA (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’ (12 February 2019) FRA Opinion 2/2019

United Nations documents

‘Letter of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism’ (7 December 2018) OL OTH 71/2018

UNCHR, ‘Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression’ (11 May 2016) UN Doc A/HRC/32/38

UNCHR ‘Report of the Special Rapporteur on Violence against Women, Its Causes and Consequences on Online Violence against Women and Girls from a Human Rights Perspective’ (18 June 2018) UN Doc A/HRC/38/47

UNCHR ‘Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression’ (9 October 2019) UN Doc A/74/486

Council of Europe documents

Council of Europe, ‘Liability and Jurisdictional Issues in Online Defamation Cases’ (September 2019) DGI (2019) 04

Council of Europe (Rapporteur Karen Yeung), ‘Responsibility and AI: A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework’ (September 2019) Study DGI (2019) 05

Ireland policy documents

Department of Communications, Climate Action and Environment, ‘General Scheme of the Online Safety & Media Regulation Bill 2019: Amending the Broadcasting Act 2009’ (January 2019) <https://bit.ly/33jN0xq> accessed 10 June 2020

Facebook Ireland Limited, ‘Public Consultation on the Regulation of Harmful Content on Online Platforms and the Implementation of the revised Audiovisual Media Services Directive: Submission of Facebook Ireland Limited’ (2019) <https://bit.ly/2PirOjb> accessed 10 June 2020

Google, ‘Public consultation on the regulation of harmful content on online platforms and the implementation of the revised Audiovisual Media Services Directive’ (2019) <https://bit.ly/2PirOjb> accessed 10 June 2020

Secondary Sources

Books and articles

Abbott KW, Levi-faur D and Snidal D, ‘Theorizing Regulatory Intermediaries: The RIT Model’ (2017) 610 The ANNALS of the American Academy of Political and Social Science 14

—— and others, ‘Competence versus Control: The Governor’s Dilemma’ (2019) Regulation & Governance 1 <https://bit.ly/2BUxVat> accessed 20 May 2020

AccessNow, ‘Two Years Under the EU GDPR: An Implementation Progress Report’ (May 2020) <https://bit.ly/33jMCyU> accessed 15 June 2020

Alexy R, *A Theory of Constitutional Rights* (Julian Rivers tr, OUP 2002)

Amalfitano C, *General principles of EU law and the Protection of Fundamental Rights* (Edward Elgar Publishing 2018)

Ananny M, ‘Making up Political People: How Social Media Create the Ideals, Definitions, and Probabilities of Political Speech’ (2020) 4 Georgetown Law Technology Review 351

—— and Crawford K, ‘Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability’ (2018) 20 New Media & Society 973

Andersen J and Søe SO, ‘Communicative Actions We Live by: The Problem with Fact-checking, Tagging or Flagging Fake News – The Case of Facebook’ (2020) 35 European Journal of Communication 126

Ayres I and Braithwaite J, *Responsive Regulation: Transcending the Deregulation Debate* (OUP 1992)

Balkin JM, ‘Deconstructive Practice and Legal Theory’ (1987) 96 Yale Law Journal 743

—— ‘Constitutional Crisis and Constitutional Rot’ in Mark A Graber, Sanford Levinson and Mark Tushnet (eds), *Constitutional Democracy in Crisis?* (OUP 2018)

—— 'Free Speech Is a Triangle' (2018) 118 Columbia Law Review 2011

Barbour C, 'Between Politics and Law: Hannah Arendt and the Subject of Rights' in Marco Goldoni and Christopher McCorkindale (eds), *Hannah Arendt and the Law* (Hart Publishing 2012)

Bassini M, 'Fundamental Rights and Private Enforcement in the Digital Age' (2019) 25 European Law Journal 182

Beijer M, *The Limits of Fundamental Rights Protection by the EU: The Scope for the Development of Positive Obligations* (Intersentia 2017)

Benkler Y, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press 2006)

—— 'Degrees of Freedom, Dimensions of Power' (2016) 145 Daedalus 18

Birch K, Chiappetta M and Artyushina A, 'The Problem *of* Innovation in Technoscientific Capitalism: Data *Rentiership* and the Policy Implications of Turning Personal Digital Data into a Private Asset' (in press, 2020) Policy Studies <https://bit.ly/3hWpOcP> accessed 23 July 2020

Birnhack MD and Elkin-Koren N, 'The Invisible Handshake: The Reemergence of the State in the Digital Environment' (2003) 8 Virginia Journal of Law and Technology 1

Black J, 'Decentring Regulation: Understanding The Role of Regulation and Self-Regulation in a 'Post-regulatory' World' (2001) 54 Current Legal Problems 103

Boonen C, 'Limits to the Politics of Subjective Rights: Reading Marx After Lefort' (2019) 30 Law & Critique 179

Britton-Purdy J and others, 'Building a Law-and-Political-Economy Framework: Beyond the Twentieth-Century Synthesis' (2020) 129 The Yale Law Journal 1784

Busch C, 'Self-regulation and Regulatory Intermediation in the Platform Economy' in Marta Cantero Gamito and Hans-W Micklitz (eds), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes* (Edward Elgar Publishing 2020)

Cafaggi F and Pistor K, 'Regulatory Capabilities: A Normative Framework for Assessing the Distributional Effects of Regulation' (2015) 9 Regulation & Governance 95

Callamard A, 'The Human Rights Obligations of Non-State Actors' in Rikke Frank Jørgensen (ed), *Human Rights in the Age of Platforms* (MIT Press 2019)

Cappello M (ed), *Self- and Co-regulation in the New AVMSD* (IRIS Special, European Audiovisual Observatory 2019)

Castells M, 'The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance' (2008) 616 The ANNALS of the American Academy of Political and Social Science 78

—— *The Rise of the Network Society* (2nd edn, Wiley-Blackwell 2010)

Citron DK, 'Extremist Speech, Compelled Conformity, and Censorship Creep' (2018) 93 Notre Dame Law Review 1035

Coche E, 'Privatised Enforcement and the Right to Freedom of Expression in a World Confronted with Terrorism Propaganda Online' (2018) 7(4) Internet Policy Review 1

Cohen JE 'Affording Fundamental Rights: A Provocation Inspired by Mireille Hildebrandt' (2017) 4 Critical Analysis of Law 78

—— *Between Truth and Power: Legal Constructions of Informational Capitalism* (OUP 2019)

—— 'Tailoring Election Regulation: The Platform Is the Frame' (2020) 4 Georgetown Law Technology Review 641

Couldry N and Kallinikos J, 'Ontology' in Jean Burgess, Alice Marwick and Thomas Poell (eds), *The SAGE Handbook of Social Media* (SAGE Publications 2017)

Dobber T, Fathaigh RÓ and Borgesius FJZ, 'The Regulation of Online Political Micro-targeting in Europe' (2019) 8(4) Internet Policy Review 1

Edwards L, ''With Great Power Comes Great Responsibility?': The Rise of Platform Liability' in Lilian Edwards (ed), *Law, Policy and the Internet* (Hart Publishing 2019)

European Union Agency for Fundamental Rights, *Freedom to Conduct a Business: Exploring the Dimensions of a Fundamental Right* (Publications Office of the European Union 2015)

Finck M, 'Digital Co-regulation: Designing a Supranational Legal Framework for the Platform Economy' (2017) LSE Legal Studies Working Paper No. 15/2017 <https://bit.ly/3fpMSPk> accessed 23 March 2020

—— 'Automated Decision-Making and Administrative Law' (2019) Max Planck Institute for Innovation & Competition Research Paper, No. 19-10 <https://bit.ly/39SVM6P> accessed 20 June 2020

Fontanelli F, 'The Court of Justice of the European Union and the Illusion of Balancing in Internet-Related Disputes' in Oreste Pollicino and Graziella Romeo (eds), *The Internet and Constitutional Law: The Protection of Fundamental Rights and Constitutional Adjudication in Europe* (Routledge 2016)

Frantziou E, *The Horizontal Effect of Fundamental Rights in the European Union: A Constitutional Analysis* (OUP 2019)

—— '(Most of) the Charter of Fundamental Rights is Horizontally Applicable' (2019) 15 European Constitutional Law Review 306

Gardbaum S, 'The Structure and Scope of Constitutional Rights' in Rosalind Dixon and Tom Ginsburg (eds), *Comparative Constitutional Law* (Edward Elgar Publishing 2011)

Gillespie T, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018)

—— 'Platforms Are Not Intermediaries' (2018) 2 Georgetown Law Technology Review 198

Gordon C, 'Governmental Rationality: An Introduction' in Graham Burchell, Colin Gordon and Peter Miller (eds), *The Foucault Effect: Studies in Governmentality: With Two Lectures by and an Interwiev with Michel Foucault* (University of Chicago Press 1991)

Gorwa R, 'The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content' (2019) 8(2) Internet Policy Review 1

—— 'What Is Platform Governance?' (2019) 22 Information, Communication & Society 854

—— Binns R and Katzenbach C, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020)(Jan–Jun) Big Data & Society 1

—— and Ash TG, 'Democratic Transparency in the Platform Society' in Nate Persily and Josh Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (forthcoming, CUP 2020) <https://osf.io/preprints/socarxiv/ehcy2/> accessed 28 May 2020

Graber CB, 'Artificial Intelligence, Affordances and Fundamental Rights' in Mireille Hildebrandt and Kieron O'Hara (eds), *Life and the Law in the Era of Data-driven Agency* (Edward Elgar Publishing 2020)

Grewal DS and Purdy J, 'Introduction: Law and Neoliberalism' (2014) 77 Law & Contemporary Problems 1

Grimmelmann J, 'The Virtues of Moderation' (2015) 17 Yale Journal of Law & Technology 42

—— 'The Platform Is the Message' (2018) 2 Georgetown Law Technology Review 217

Helberger N, 'The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power' (2020) Digital Journalism <https://bit.ly/3hY3akg> accessed 27 July 2020.

Hildebrandt M, *Smart Technologies and the End(s) of Law* (Edward Elgar Publishing 2015)

Hoffmann AL, Proferes N and Zimmer M, '"Making the World More Open and Connected": Mark Zuckerberg and the Discursive Construction of Facebook and Its Users' (2018) 20 New Media & Society 199

Hofmann J, Katzenbach C and Gollatz K, 'Between Coordination and Regulation: Finding the Governance in Internet Governance' (2017) 19 New Media & Society 1406

Ingram JD, 'What Is a “Right to Have Rights”? Three Images of the Politics of Human Rights' (2008) 102 American Political Science Review 401

Jørgensen RF, 'Framing Human Rights: Exploring Storytelling within Internet Companies' (2018) 21 Information, Communication & Society 340

—— ‘Rights Talk: In the Kingdom of Online Giants’ in Rikke Frank Jørgensen (ed), *Human Rights in the Age of Platforms* (MIT Press 2019)

—— ‘When Private Actors Govern Human Rights’ in Ben Wagner, Mattias C Kettemann and Kilian Vieth, *Human Rights and Digital Technology* (Edward Elgar Publishing 2019)

—— and Møller Pedersen A, ‘Online Service Providers as Human Rights Arbiters’ in Mariarosaria Taddeo and Luciano Floridi (eds), *The Responsibilities of Online Service Providers* (Springer 2017)

—— and Zuleta L, 'Private Governance of Freedom of Expression on Social Media Platforms' (2020) 41(1) Nordicom 51

Kapczynski A, ‘The Law of Informational Capitalism’ (Book review) (2020) 129 The Yale Law Journal 1460

Kaye D, *Speech police: The Global Struggle to Govern the Internet* (Columbia Global Reports 2019)

Keller D, ‘Internet Platforms: Observations on Speech, Danger, and Money’ (2018) Hoover Institution Essay, Aegis Series Paper No. 1807 <https://bit.ly/31dLrhM> accessed 20 June 2020

—— ‘Who Do You Sue?: State and Platform Hybrid Power over Online Speech’ (2019) Hoover Institution Essay, Aegis Series Paper No. 1902 <https://hvr.co/39PBzz1> accessed 20 March 2020

—— and Leerssen P, ‘Facts and Where Find Them: Empirical Research on Internet Platforms and Content Moderation’ in Nate Persily and Josh Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (forthcoming, CUP 2020) <https://bit.ly/2EJAOfh> accessed 16 July 2020

Khoury S and White D, *Corporate Human Rights Violations: Global Prospects for Legal Action* (Routledge 2017)

Klonick K, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 Harvard Law Review 1598

Korpisaari P, ’Sananvapaus verkossa: yksilöön kohdistuva vihapuhe ja verkkoalustan ylläpitäjän vastuu’ (2019) Lakimies 928

Koulu R, *Law, Technology and Dispute Resolution: Privatisation of Coercion* (Routledge 2019)

Kuczerawy A, *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Safeguards* (Intersentia 2018)

—— 'The Proposed Regulation on Preventing the Dissemination of Terrorist Content Online: Safeguards and Risks for Freedom of Expression' *Center for Democracy and Technology* (5 December 2018) 9 <https://tinyurl.com/y4mr7wyj> accessed 15 March 2020

Laidlaw EB, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (CUP 2015)

Land MK, 'Regulating Private Harms Online: Content Regulation under Human Rights Law' in Rikke Frank Jørgensen (ed), *Human Rights in the Age of Platforms* (MIT Press 2019)

Lavrysen L, *Human Rights in a Positive State: Rethinking the Relationship between Positive and Negative Obligations under the European Convention on Human Rights* (Intersentia 2016)

Leerssen P, 'Cut Out by the Middle Man: The Free Speech Implications of Social Network Blocking and Banning in the EU' (2015) 6 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 99

—— 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (SSRN 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3544009> accessed 17 July 2020

Leino P, 'Secrecy, Efficiency, Transparency in EU Negotiations: Conflicting Paradigms?' (2017) 5 Politics and Governance 6

Liscow Z, 'Is Efficiency Biased?' (2018) 85 The University of Chicago Law Review 1649

Lynskey O, 'Regulating "Platform Power"' (2017) LSE Law, Society and Economy Working Papers 1/2017, 28 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2921021> accessed 25 May 2020

Maréchal N and Roberts Biddle E, 'It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge' *Ranking Digital Rights* (17 March 2020)

—— MacKinnon R and Dheere J, 'Getting to the Source of Infodemics: It's the Business Model' *Ranking Digital Rights* (26 May 2020)

Marsden CT, *Internet Co-regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (CUP 2011)

—— 'Prosumer Law and Network Platform Regulation: The Long View Towards Creating Offdata' (2018) 2 Georgetown Law Technology Review 376

Maroni M, ‘The Liability of Internet Intermediaries and the European Court of Human Rights’ in Tuomas Ojanen and Bilyana Petkova (eds), *Fundamental Rights Protection Online: The Future Regulation Of Intermediaries* (forthcoming, Edward Elgar Publishing 2020)

Mir JB and Bassini M, ‘Freedom of Expression in the Internet: Main Trends of the Case Law of the European Court of Human Rights’ in Oreste Pollicino and Graziella Romeo (eds), *The Internet and Constitutional Law: The Protection of Fundamental Rights and Constitutional Adjudication in Europe* (Routledge 2016)

Morgan B and Yeung K, *An Introduction to Law and Regulation: Text and Materials* (CUP 2007)

Morozov E, *To Save Everything, Click Here: The Folly of Technological Solutionism* (Public Affairs 2013)

Moyn S, *Not Enough: Human Rights in an Unequal World* (Harvard University Press 2018)

Naarttijärvi M, ‘Legality and Democratic Deliberation in Black Box Policing’ (2019) Technology and Regulation 35

Nieborg DB and Helmond A, ‘The Political Economy of Facebook’s Platformization in the Mobile Ecosystem: Facebook Messenger as a Platform Instance’ (2019) 41 Media, Culture & Society 196

Ohm P, ‘Regulating at Scale’ (2018) 2 Georgetown Law Technology Review 546

Palmer S, 'Privatization and Human Rights in the United Kingdom' in Tsvi Kahana and Anat Scolnicov (eds), *Boundaries of State, Boundaries of Rights* (CUP 2016)

Pariser E, *The Filter Bubble: What the Internet is Hiding from You* (Penguin 2012)

Pasquale F, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015)

—— ‘The Automated Public Sphere’ in Ann Rudinow Sætnan, Ingrid Schneider and Nicola Green (eds), *The Politics and Policies of Big Data: Big Data, Big Brother?* (Routledge 2018)

Petre C, Duffy BE and Hund E, ‘“Gaming the System”: Platform Paternalism and the Politics of Algorithmic Visibility’ (2019)(October–December) Social Media + Society 1

Pollicino O, ‘Metaphors and Judicial Frame: Why Legal Imagination (Also) Matters in the Protection of Fundamental Rights in the Digital Age’ in Bilyana Petkova and Tuomas Ojanen (eds), *Fundamental Rights Protection Online* (forthcoming, Edward Elgar Publishing 2020)

Quintel T and Ullrich C, ‘Self-Regulation of Fundamental Rights?: The EU Code of Conduct on Hate Speech, Related Initiatives and Beyond’ in Tuomas Ojanen and Bilyana Petkova (eds), *Fundamental Rights Protection Online: The Future Regulation Of Intermediaries* (forthcoming, Edward Elgar Publishing 2020)

Raymond M and DeNardis L, 'Multistakeholderism: Anatomy of an Inchoate Global Institution' (2015) 7 International Theory 572

Reed C and Murray A, *Rethinking the Jurisprudence of Cyberspace* (Edward Elgar Publishing 2018)

Roberts A, ‘Privacy, Data Retention and Domination: Digital Rights Ireland Ltd v Minister for Communications’ (2015) 78 Modern Law Review 535

Rubinstein IS, 'The Future of Self-regulation Is Co-regulation' in Evan Selinger, Jules Polonetsky and Omer Tene (eds), *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018)

Sander B, ‘Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation’ (2020) 43 Fordham Internal Law Journal 939

Sarfaty GA, ‘Measuring Corporate Accountability Through Global Indicators’ in Sally Engle Merry, Kevin E Davis and Benedict Kingsbury (eds), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (CUP 2015)

Savin A, ‘Regulating Internet Platforms in the EU: The Emergence of the ‘Level Playing Field’’ (2018) 34 Computer Law & Security Review 1215

Schneider I, ‘Bringing the State Back in: Big Data-based Capitalism, Disruption, and Novel Regulatory Approaches in Europe’ in Ann Rudinow Sætnan, Ingrid Schneider and Nicola Green (eds), *The Politics and Policies of Big Data: Big Data, Big Brother?* (Routledge 2018)

Schwarz O, ‘Facebook Rules: Structures of Governance in Digital Capitalism and the Control of Generalized Social Capital’ (2019) 36 Theory, Culture and Society 117

Schwemer SF, ‘Trusted Notifiers and the Privatization of Online Enforcement’ (2019) 35(6) Computer Law & Security Review 1

Short J, ‘Paranoid Style in Regulatory Reform’ (2011) 63 Hastings Law Journal 633

—— Toffel MW and Hugil AR, ‘Monitoring Global Supply Chains’ (2016) 37 Strategic Management Journal 1878

Sunstein CR and Thaler RH, *Nudge: Improving Decisions About Health, Wealth and Happiness* (Revised edn, Penguin 2012)

Susi M, 'The Internet Balancing Formula' (2019) 25 European Law Journal 198

Susser D, Roessler B and Nissenbaum H, ‘Online Manipulation: Hidden Influences in a Digital World’ (2019) 4(1) Georgetown Law Technology Review 1

Suzor NP, *Lawless: The Secret Rules That Govern Our Digital Lives* (CUP 2019)

—— Seignior B and Singleton J, 'Non-consensual Porn and the Responsibilities of Online Intermediaries' (2017) 40 Melbourne University Law Review 1057

—— and others, 'Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online' (2019) 11(1) Policy & Internet 84

Taylor E, *The Privatization of Human Rights: Illusions of Consent, Automation and Neutrality* (Centre for International Governance Innovation 2016)

Urban JM, Karaganis J and Schofield B, 'Notice and Takedown in Everyday Practice' (Version 2, 2016, updated March 2017) UC Berkeley Public Law Research Paper No 2755628 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628> accessed 22 June 2020

van der Sloot B, 'A New Approach to the Right to Privacy, or How the European Court of Human Rights Embraced the Non-domination Principle' (2018) 34 Computer Law & Security Review 539

van Dijck J, Poell T and de Waal M, *The Platform Society: Public Values in a Connective World* (OUP 2018)

van Gestel R and van Lochem P, 'Private Standards as a Replacement for Public Lawmaking?' in Marta Cantero Gamito and Hans-W Micklitz (eds), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes* (Edward Elgar Publishing 2020)

Vieth K, 'Policing 'Online Radicalization': The Framing of Europol's Internet Referral Unit' in Ben Wagner, Matthias C Kettemann and Kilian Vieth (eds), *Research Handbook on Human Rights and Digital Technology: Global Politics, Law and International Relations* (Edward Elgar Publishing 2019)

Wagner B, *Global Free Expression: Governing the Boundaries of Internet Content* (Springer 2016)

—— 'Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems' (2019) 11 Policy & Internet 104

Waldman AE, 'Power, Process, and Automated Decision-making' (2019) 88 Fordham Law Review 613

Ward A, 'Remedies under the EU Charter of Fundamental Rights' in Sionaidh Douglas-Scott and Nicholas Hatzis (eds), *Research Handbook on EU Law and Human Rights* (Edward Elgar Publishing 2017)

Williams A, 'Human Rights in the EU' in Damian Chalmers and Anthony Arnull (eds), *The Oxford Handbook of European Union Law* (OUP 2015)

Woods L, 'Social Media: It Is Not Just about Article 10' in David Mangan and Lorna E Gillies (eds), *The Legal Challenges of Social Media* (Edward Elgar Publishing 2017)

Xenos D, *The Positive Obligations of the State under the European Convention of Human Rights* (Routledge 2012)

Yeung K, ‘Better Regulation, Administrative Sanctions and Constitutional Values’ (2013) 33 Legal Studies 312

—— ‘‘Hypernudge’: Big Data as a Mode of Regulation by Design’ (2017) 20 Information, Communication & Society 118

—— ‘Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism’ (2018) 8 International Data Privacy Law 258

Zuboff S, *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power* (Profile Books 2019)

News articles

‘Plans for ‘first-of-its-kind’ online safety regulator published’ *Irish Legal News* (10 January 2020) <https://bit.ly/3k5A0kY> accessed 10 June 2020

Ackerman E and others, ‘A letter on justice and open debate’ *Harper’s Magazine* (7 July 2020) <https://harpers.org/a-letter-on-justice-and-open-debate/> accessed 19 July 2020

Biddle S, Ribeiro PV and Dias T, ‘Invisible Censorship’ *The Intercept* (16 March 2020) <https://tinyurl.com/y4eolz7o> accessed 3 August 2020.

Chen A, ‘The laborers who keep dick pics and beheadings out of your Facebook feed’ *WIRED* (23 October 2014) <www.wired.com/2014/10/content-moderation/> accessed 16 July 2020

Diresta R, ‘Free speech is not the same as free reach’ *WIRED* (20 October 2018) <www.wired.com/story/free-speech-is-not-the-same-as-free-reach/> accessed 19 July 2020

Gorey C, ‘Online safety law could block social media sites for non-compliance’ *Siliconrepublic* (10 Jan 2020) <www.siliconrepublic.com/companies/online-safety-law-ireland-details> accessed 10 June 2020

Halliday J, ‘Twitter's Tony Wang: ‘We are the free speech wing of the free speech party’’ *The Guardian* (22 March 2012) <https://bit.ly/3gnKf1I> accessed 22 May 2020

Horwitz J and Seetharaman D, ‘Facebook executives shut down efforts to make the site less divisive’ *Washington Post* (26 May 2020) <https://on.wsj.com/2PdX4js> accessed 23 July 2020

Lewis H, ‘How capitalism drives cancel culture’ (14 July 2020) <https://bit.ly/39O2Z8h> accessed 19 July 2020

Luckerson V, ‘Here’s how Facebook’s news feed actually works’ *TIME* (9 July 2015) <https://time.com/collection-post/3950525/facebook-news-feed-algorithm/> accessed 19 July 2020

McCarthy T, ‘Zuckerberg says Facebook won't be 'arbiters of truth' after Trump threat’ *The Guardian* (28 May 2020) <https://bit.ly/3gnnTO6> accessed 16 July 2020

Meyer R, ‘When a snuff film becomes unavoidable’ *The Atlantic* (26 August 2015) <https://bit.ly/3fm5D6n> accessed 6 June 2020

Murphy, M and others, ‘Tech results: Amazon, Apple and Facebook earn record profits during pandemic’ *The Telegraph* (20 July 2020) <https://bit.ly/2DrFt4Y> accessed 2 August 2020

Ranking J, ‘EU says China behind 'huge wave' of Covid-19 disinformation’ *The Guardian* (10 June 2020) <https://bit.ly/3hXZbUQ> accessed 2 August 2020

Solon O, ‘Twitter launches another bid to tackle bots and abuse after years of promises’ *The Guardian* (1 March 2018) <https://bit.ly/30jtzCY> accessed 22 May 2020

Solon O, ‘Facebook ignored racial bias research, employees say’ *NBC News* (23 July 2020) <https://nbcnews.to/3fg9cLm> accessed 29 July 2020.

Wintour P, ‘Mark Zuckerberg: Facebook must accept some state regulation’ *The Guardian* (15 February 2020) <https://bit.ly/33pQ3nI> accessed 14 April 2020

Wong JC, ‘Twitter to ban all political advertising, raising pressure on Facebook’ *The Guardian* (20 October 2019) <https://bit.ly/39Qgnc2> accessed 19 July 2020

Blog posts

Ananny M, ‘Probably speech, maybe free: toward a probabilistic understanding of online expression and platform governance’ *Knight First Amendment Institute* (21 August 2019) <https://bit.ly/31ckHhW> (accessed 19 July 2020)

Bickert M and Fishman B, ‘Hard Questions: What are we doing to stay ahead of terrorists?’ *Facebook Newsroom* (8 November 2018) <https://bit.ly/3gyq8hm> accessed 17 July 2020

Butler C, ‘Official EU agencies falsely report more than 550 Archive.org URLs as terrorist content’ *Internet Archive* (10 April 2020) <https://bit.ly/2EFE0Zf> accessed 26 June 2020

Gleicher N, ‘Removing more coordinated inauthentic behavior from Iran’ *Facebook Newsroom* (28 May 2019) <https://about.fb.com/news/2019/05/removing-more-cib-from-iran/> accessed 23 July 2020

Jeppesen J-H, Nosák D and Esposito P, ‘EU Tech Policy Brief: March 2020 Recap’ *Center for Democracy & Technology* (3 April 2020) <https://bit.ly/3gyq3KA> accessed 4 May 2020

Keller D, ‘Broad Consequences of a Systemic Duty of Care for Platforms’ *The Center for Internet and Society and Stanford Law School* (1 June 2020) <https://stanford.io/2XkFFtG> accessed 29 June 2020

Romero J, ‘Taking legal action against those who abuse our services’ *Facebook Newsroom* (18 June 2020) <https://about.fb.com/news/2020/06/automation-software-lawsuits/> accessed 19 July 2020.

Websites

EDiMA Website, <https://edima-eu.org/> accessed 8 June 2020

EDRi, ‘EU Internet Forum against terrorist content and hate speech online: Document pool’ (10 March 2016) <https://edri.org/eu-internet-forum-document-pool/> accessed 22 May 2020

EDRi, ‘EDRi and Access Now withdraw from the EU Commission IT Forum discussions’ (31 May 2016) <https://bit.ly/2EJ7N3m> accessed 23 July 2020

Department of Communications, Climate Action and Environment, ‘Minister Bruton Publishes Draft Scheme of New Online Safety Law’ (10 January 2019) <https://bit.ly/3hZ7505> accessed 10 June 2020

Department of Communications, Climate Action and Environment, ‘Regulation of Harmful Online Content and the Implementation of the revised Audiovisual Media Services Directive’ (April 2019) <https://bit.ly/3grnZEk> accessed 10 June 2020

Europol, ‘EU Internet Referral Unit - EU IRU: Monitoring terrorism online’ (2020) <www.europol.europa.eu/about-europol/eu-internet-referal-unit-eu-iru> accessed 15 May 2020

Facebook, ‘Community Standards’ *Facebook Inc* (2020) <www.facebook.com/communitystandards/> (accessed 25 February 2020)

Facebook, ‘Data Policy’ *Facebook Inc* (19 April 2018) <www.facebook.com/about/privacy> accessed 24 February 2020

Facebook Inc, ‘Facebook Reports Fourth Quarter and Full Year 2019 Results’ (29 January 2020) <https://bit.ly/3i5FCdr> accessed 14 April 2020

Facebook, ‘Product Policy Forum Minutes’ *Facebook Inc* (15 November 2018) <https://bit.ly/2EFDnPn> accessed 22 May 2020

Facebook Transparency, ‘Content Restrictions Based on Local Law: Jan – Jun 2019’ *Facebook Inc* <https://transparency.facebook.com/content-restrictions> accessed 17 April 2020

OnlineCensorship.org, ’How to Appeal’ <https://tinyurl.com/y48x9ezr> accessed 22 July 2020

‘Overview’ *Amazon Mechanical Turk Inc* <www.mturk.com/> accessed 2 August 2020

PwC Ireland, '7 things you need to know about the proposed Online Safety and Media Regulation Bill' (29 January 2020) <https://pwc.to/2PeVu0N> accessed 6 June 2020

Simonovic D and Kaye D, 'UN experts urge States and companies to address online gender-based abuse but warn against censorship' *OHCHR* (8 March 2017) <https://bit.ly/39PADuv> accessed 5 March 2020

Twitter, 'Defending and respecting the rights of people using our service' *Twitter Inc* (2020) <https://bit.ly/33j7syH> accessed 17 April 2020

Twitter, 'Our Advocacy' *Twitter Inc* (2020) <https://tinyurl.com/y256pwwa> accessed 7 March 2020

Twitter, 'Safety at Twitter' *Twitter, Inc* (2020) <https://about.twitter.com/en_us/safety/safety-partners.html> accessed 15 March 2020

Twitter Transparency Report, 'Removal Requests – January to June 2019: No Action' *Twitter Inc* (2020) <https://transparency.twitter.com/en/removal-requests.html> accessed 17 April 2020

YouTube, 'About' *YouTube LLC* (2020) <www.youtube.com/intl/en-GB/about/> accessed 7 March 2020.

YouTube, 'YouTube for the Press' *YouTube LLC* <www.youtube.com/about/press/> accessed 14 April 2

Abbreviations

CJEU	Court of Justice of the European Union
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EU IRU	Europol Internet Referral Unit
FRA	European Union Agency for Fundamental Rights
TFEU	Treaty on the Functioning of the European Union

1. Introduction

When thinking about the subject of this thesis, I started, as I believe one often starts with legal research, from a concern over an injustice in society that I had perceived and which for some reason had stuck to my mind. There is likely a host of explanations to why one perceives some societal injustices and not others or, why some perceived injustices stuck while others are ephemeral. The concern over an injustice stuck to my mind departs from the now so often repeated statement that it has probably already flattened to a banality. It is that now a handful of multinational corporations intermediate and disintermediate communication among people on network settings that have become to be labeled as 'social media'. These are namely Facebook, YouTube, Twitter, and perhaps a couple others.

I had understood that intermediation and disintermediation by Facebook and others would afford substantial influence on communication. Moreover, intermediation and disintermediation are not accidental but purposeful. I had also remarked the increasingly vocal concerns over how the intermediation and disintermediation by corporations implicate people's wellbeing and produce myriad other harms that had proliferated on these networks, ranging from racial hatred and sexual harassment to election manipulation. However, more dauntingly I noticed that many had little understanding of why and how these harms had come into being (perhaps because I didn't understand myself either) or little control to react to them.

Yet I had taken note of the flurry of relevant initiatives set in motion by the EU. Some initiatives had already been implemented. They were imposing regulation – it was something. What I was finding troubling was that the regulation introduced for rights protection didn't seem to interrogate the private power of (dis)intermediation, but rather it appeared to facilitate and cement it. The EU seemed to deeply involve the corporations in its own regulatory efforts. I started to look for a piece of fundamental rights doctrine that could serve as a backdrop against the change I thought I had observed. While there are undoubtedly other viable approaches that I could have opted, the ECtHR doctrine on state's positive obligations resonated particularly well for me. Effectively, the positive obligations require a state to put in place an adequate legislative and administrative framework to prevent and remedy violations to fundamental rights, whether inflicted by public authorities or private actors. The conundrum, however, was that this was exactly what the EU seemed to be doing: implementing a regulatory and administrative framework for addressing harms related to these services. But was it 'adequate'? Therefore, I ended up situating my thesis at the intersection of this fundamental rights 'regulation mandate' and the actual regulatory practice incorporating the new trend of private involvement.

1.1 Scope of Analysis, Research Questions, and Definitions

1.1.1 Scope and Research Questions

This thesis explores how the EU protects/implicates freedom of expression and information by regulating the user-generated content and adverts on social media, and how this regulation distributes power. One should be mindful that freedom of expression is hardly the only fundamental right implicated on social media but instead

a wide array of rights is at stake, ranging from non-discrimination to privacy to rights of the child and so on. However, I am approaching the issues on social media from the perspective of freedom of expression. Yet one must be wary here. The conventional truisms of freedom of expression and information may promote ideals and doomsday visions that might from time to time need rethinking. I am trying to avoid locking into such uncontested axioms too strictly. My point of departure for the argumentation in this thesis is that the requirements of freedom of expression are contextual and thus often far from self-evident, especially in the still relatively unexplored setting of social media. In addition, many issues online lend themselves to be assessed through free expression as well as with the lenses of other rights.[1] Rights don't always produce contradictory demands but are often concurrent and interdependent.[2] Thus, it may at times be appropriate to involve other rights too, especially where they seem to connect with freedom of expression. However, data extraction on the Internet, for instance, is not central in this thesis. On top of this basis, I am putting forward four delineating points that draw the scope of analysis further.

Firstly, my interest focuses on the degree of involvement of private actors, most prominently the social media companies, in the regulation of people's freedom of expression. To be sure, remarks on 'private enforcement'[3] and 'privatization of human rights'[4] have become somewhat commonplace in digital rights scholarship. In my view, however, those remarks have often produced slightly isolated criticisms without a systematic analysis of the broader framework through which the 'privatization' is to be actuated. From this it follows that my thesis concerns primarily the *structure* of freedom of expression on social media. The structure of rights is not widely held to be an independent sub-field of rights study.[5] However, according to Gardbaum: 'The structure of constitutional rights can be usefully distinguished from their substance. (…) [T]he structure is the underlying framework – set of concepts, principles, doctrines and institutions – that applies to, organizes and characterizes as a whole within that system'.[6] The structure of rights as a sub-field of doctrine encompasses questions on the limitations of rights, vertical and horizontal effect of rights, and negative and positive rights.[7] The specific focus in my thesis are positive and negative rights obligations, which are closely related to the themes of vertical and horizontal effect.[8] While rights scholarship often focuses on interrogating specific laws' substance, I believe my shift of attention to structure allows me zoom out to see the potential changes in the broader

[1] Lorna Woods, 'Social Media: It Is Not Just about Article 10' in David Mangan and Lorna E Gillies (eds), *The Legal Challenges of Social Media* (Edward Elgar Publishing 2017) 115–124.
[2] Emily Taylor, *The Privatization of Human Rights: Illusions of Consent, Automation and Neutrality* (Centre for International Governance Innovation 2016) 2.
[3] Marco Bassini, 'Fundamental Rights and Private Enforcement in the Digital Age' (2019) 25 European Law Journal 182; and Eugénie Coche, 'Privatised Enforcement and the Right to Freedom of Expression in a World Confronted with Terrorism Propaganda Online' (2018) 7(4) Internet Policy Review 1.
[4] Taylor (n 2).
[5] Stephen Gardbaum, 'The Structure and Scope of Constitutional Rights' in Rosalind Dixon and Tom Ginsburg (eds), *Comparative Constitutional Law* (Edward Elgar Publishing 2011) 387.
[6] ibid.
[7] ibid 388, 391–392, 396–397.
[8] Eleni Frantziou, *The Horizontal Effect of Fundamental Rights in the European Union: A Constitutional Analysis* (OUP 2019) 39–40;

picture. Yet focusing on the structure does not mean complete detachment from the substance.[9] Addressing the structure in complete isolation could be counterproductive as without any outline of the substantive implications to freedom of expression, it might be hard to concretize the structure as well.

Secondly, without prejudice to the notion of acontextual indeterminacy I remarked above, some legal reference points are of course necessary. My principal one here is the European Convention on Human Rights[10] (hereafter the ECHR) as 'the flagship treaty' for European-wide human rights codification.[11] In addition, the structural typology of positive and negative fundamental right obligations has also developed in the jurisprudence of the European Court of Human Rights (hereafter the ECtHR) as the authoritative interpreter of the ECHR. However, because the aim is to analyze EU regulation, and because the European-level rights system has become pluralistic, it is not possible to exclude the Charter of the Fundamental Rights of the EU (hereafter the Charter).[12] Moreover, it is clear that in general the most relevant articles for freedom of expression and information are Article 10 of the ECHR and Article 11 of the Charter.

Thirdly, as indicated in the title, my primary interest is targeted to the companies that operate a social media service. However, I am introducing also a limitation here: I limit my analysis primarily to the largest services, i.e. social networking service Facebook (including Instagram as owned by the former), micro-blogging service Twitter and video-sharing platform YouTube (as owned by Google/Alphabet).[13] The focus on the biggest social media services may at first appear rather arbitrary but, in my view, it receives backing from a branch of the Internet regulation studies which are concerned on the so-called 'Internet gatekeepers'.[14] Namely, I am referencing to an influential study by Emily B Laidlaw, who identifies a wide array of different Internet gatekeepers which exercise control over information flows by 'channeling information through a channel, deleting information or shaping information into a particular form'.[15] A gatekeeper, 'as a result of this control of the flow of information, facilitate or hinder deliberation and participation in democratic culture'.[16] Indeed, social media platforms are conveniently placed at the intersection of massive information flows of users' social interaction: information flows through their platforms.[17] Laidlaw's point is that gatekeeping is a relevant use of power from the fundamental rights perspective, and while also small Internet businesses and even

[9] Gardbaum (n 5) 402.
[10] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended).
[11] Joan Barata Mir and Marco Bassini, 'Freedom of Expression in the Internet: Main Trends of the Case Law of the European Court of Human Rights' in Oreste Pollicino and Graziella Romeo (eds), *The Internet and Constitutional Law: The Protection of Fundamental Rights and Constitutional Adjudication in Europe* (Routledge 2016) 71.
[12] Charter of Fundamental Rights of the European Union [2012] OJ C326/391.
[13] In 2016, a study by the Commission identified Facebook, YouTube, Twitter and Instagram as the four most popular platforms in terms of online traffic in Europe and (worldwide) monthly active users. European Commission, 'Behavioural Study on Advertising and Marketing Practices in Online Social Media' (Final Report, June 2018) 14–16.
[14] Chris Reed and Andrew Murray, *Rethinking the Jurisprudence of Cyberspace* (Edward Elgar Publishing 2018) 159–163.
[15] Emily B Laidlaw, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (CUP 2015) 231.
[16] ibid 57.
[17] Julie E Cohen, *Between Truth and Power: Legal Constructions of Informational Capitalism* (OUP 2019) 219–220.

individuals may act as gatekeepers, their control is far more limited compared to those gatekeepers that control the vital access points in the whole Internet structure.[18] According to Laidlaw, the gatekeepers fundamental rights obligations should correspond their gatekeeping power, with the most powerful gatekeepers having the most stringent constraining obligations as well.[19] I agree with Laidlaw and I believe understanding social media service providers as vital Internet gatekeepers justifies my limitation made above.[20]

Fourthly, I am pointing out that my thesis does not focus on analyzing the rules of perhaps the most fundamental piece of social media content regulation, namely the EU e-Commerce Directive.[21] As the Directive came into force in 2000, a lot of insightful research has already been made on its role in defining the responsibilities of social media companies.[22] Instead, this thesis limits it scope to the more recent regulatory additions introduced during the last five years of the Juncker Commission, even though the most relevant rules of the Directive are briefly revisited throughout. The more recent regulation builds on top of the framework laid out by e-Commerce Directive. Based on this outline, I believe I am able to formulate the following research questions: Who has the power over people's freedom of expression on social media, including the power to define the meaning of the right itself? Through what regulatory techniques EU social media content regulation allocates this power? How the power allocation relates to the dominant fundamental rights doctrine? All three questions are intertwined and thus the answers to them should produce a comprehensive understanding on the involvement of private actors to fundamental rights protection on social media and its power implications.

1.1.2 Definitions

As a start, regarding the term 'freedom of expression', I must underline that while using the term, I mean freedom of expression and information as it is stated in Article 10 of ECHR and Article 11 of the Charter. For the sake of brevity and ease of language, I also use term 'free expression' throughout the thesis. Next, my definition for the somewhat ambiguous term 'regulation' is functional, narrow and practically aligns with a legal system. I use the term regulation and law interchangeably. Regulation denotes 'deliberate attempts by the state [or the EU] to influence socially valuable behavior which may have adverse side-effects by establishing, monitoring and enforcing legal rules'.[23] The tripartite classification means that there needs to be an element of norm-setting (establishment of rules), a way to gather information on the operation of the system (monitoring and sometimes, more narrowly, 'policing'), and a capacity to modify behavior to sufficiently conform with norms (enforcement).[24] In turn, I am adopting the term 'governance' to encompass the different

[18] Laidlaw (n 15) 47–56. The Internet infrastructure offers also other numerous access points for power, see Yochai Benkler, 'Degrees of Freedom, Dimensions of Power' (2016) 145 Daedalus 18.
[19] Laidlaw (n 15) 45–46, 243.
[20] This of course does not presuppose that small gatekeepers would harm-free or should not be regulated.
[21] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') [2000] OJ L178/1.
[22] One of the most in-depth inquiry is perhaps Aleksandra Kuczerawy, *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Safeguards* (Intersentia 2018).
[23] Bronwen Morgan and Karen Yeung, *An Introduction to Law and Regulation: Text and Materials* (CUP 2007) 3.
[24] Reed and Murray (n 14) 140.

forms of control more widely, including both institutional and social control. Governance presupposes a conflict between (implicit) expectations of at least two people.[25] It is Foucault's 'conduct of conduct', that is, 'a form of activity aiming to shape, guide or affect the conduct of some person or persons'.[26]

Governance also encompasses self-regulation 'where business or industry sectors formulate codes of conduct or operating constraints on their own initiative for which they are responsible for enforcing'.[27] The even more elusive 'co-regulation' is 'a mechanism whereby the Union Legislator entrusts the attainment of specific policy objectives set out in legislation or other policy documents to parties which are recognized in the field (such as economic operators, social partners, nongovernmental organisations, standardisation bodies or associations)'.[28] It is meant to combine the benefits of legally binding regulation and more flexible self-regulation.[29] However, it should be stressed already here that the border between 'self' and 'co' is fluid and contested.

Following the regulatory concepts, I am defining the crucial term of (social media) 'platform'. Confusion is often created when a platform refers to both the Internet company itself and as the service they offer.[30] For instance, Facebook can refer to the social media platform site or the company Facebook Inc operating the site. My platform definition is not legalistic, even though I briefly address it in Chapter 2.2. While the question of the legal categorization of platforms is important, it tends to sideline some political-economic underpinnings in a platform's purpose and functioning that play an important role in this thesis. Instead, following Gillespie, social media platforms are services that 'a) host, organize, and circulate users' shared content or social interactions for them, b) without having produced or commissioned (the bulk of) that content, c) built on an infrastructure, beneath that circulation of information, for processing data for customer service, advertising, and profit'.[31] By contrast, following the example of Gorwa, I use the term 'platform company' to refer to a multinational corporation operating a social media platform.[32]

Lastly, following Grimmelmann, the term 'content moderation' denotes 'the governance mechanisms that structure participation' on social media platforms.[33] Its sanction is social exclusion.[34] However, I need to make a couple of additions to this definition. Content moderation in this thesis are the governance mechanisms

[25] Jeanette Hofmann, Christian Katzenbach and Kirsten Gollatz, 'Between Coordination and Regulation: Finding the Governance in Internet Governance' (2017) 19 New Media & Society 1406, 1418.
[26] Colin Gordon, 'Governmental Rationality: An Introduction' in Graham Burchell, Colin Gordon and Peter Miller (eds), *The Foucault Effect: Studies in Governmentality: With Two Lectures by and an Interwiev with Michel Foucault* (University of Chicago Press 1991) 2.
[27] European Commission, 'Better Regulation Toolbox: TOOL #18 – The choice of policy instruments' 109 <https://ec.europa.eu/info/files/better-regulation-toolbox-18_en> accessed 18 June 2020.
[28] ibid.
[29] ibid 109–110.
[30] Robert Gorwa, 'What Is Platform Governance?' (2019) 22 Information, Communication & Society 854, 856.
[31] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018) 18.
[32] Gorwa (n 30) 856.
[33] James Grimmelmann, 'The Virtues of Moderation' (2015) 17 Yale Journal of Law & Technology 42, 47.
[34] Ori Schwarz, 'Facebook Rules: Structures of Governance in Digital Capitalism and the Control of Generalized Social Capital' (2019) 36 Theory, Culture and Society 117, 131.

through which the platform companies structure participation on social media platforms *to further their political-economic interests.* Of course, it is possible that content moderation advances other interests at the same time, including fundamental rights. Yet the root motive of content moderation as a corporate activity is to advance the interests of companies.[35] Content moderation is also gatekeeping comprising several gatekeeping strategies.[36] Inspired by Gorwa and others, I am dividing content moderation further to 'hard moderation' and 'soft moderation'.[37] Hard moderation is disciplinary and soft moderation non-disciplinary governance. In turn, 'content regulation' denotes the deliberate attempts by the state or the EU to influence content moderation through legal norms. Self-/co-regulation may target content moderation as well. Content moderation is addressed at length in Chapter 2.

1.2 Method and Structure

As I deem it advisable to proclaim a methodological approach as formulated in literature, this thesis takes up one that is termed as 'law and political economy approach'.[38] The approach sees law as an inextricable part of political economy as law constructs the facilitative conduits between economic and political power, not least because law creates the market.[39] It is not, therefore, an interdisciplinary approach in a sense that it would imagine three isolated realms of politics, economy and law.[40] It is also a critical approach with an explicit normative aim to advance democracy.[41] More specifically, I am drawing upon Julia E Cohen's analysis of power in the 'regulatory state' in her book *Between Truth and Power: Legal Constructions of Informational Capitalism*.[42] According to Cohen, the emergent political economy of 'informational capitalism' requires a counterpart in social and political ideology for legitimizing and facilitating economic activity.[43] This ideology is neoliberal governmentality, which, like classical liberal governmentality before it, promotes the primacy of private ordering of the market. However, (classical) liberal governmentality prescribed also robust state stewardship in the market to account for its tendency toward 'monopoly, destructive extraction, and rent-seeking'.[44] The difference that comes with the prefix 'neo' is that '[n]eoliberal governmentality does not simply

[35] ibid 121–122.
[36] Laidlaw (n 15) 45.
[37] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020)(Jan–Jun) Big Data & Society 1, 3.
[38] Jedediah Britton-Purdy and others, 'Building a Law-and-Political-Economy Framework: Beyond the Twentieth-Century Synthesis' (2020) 129 The Yale Law Journal 1784, 1784.
[39] Amy Kapczynski, 'The Law of Informational Capitalism' (Book review) (2020) 129 The Yale Law Journal 1460, 1481–1482.
[40] Britton-Purdy and others (n 38) 1792–1794, 1820.
[41] ibid 1827–1832. As regards the meaning of democracy, 'law should shape the economy to support the institutions and capacities that uphold the equality and efficacy of democratic citizens' (1830).
[42] Cohen (n 17).
[43] For Cohen, who builds her analysis on the theory of Manuel Castells, informational capitalism means: 'Capitalism " [which] is oriented toward profit-maximizing, that is, toward increasing the amount of surplus appropriated by capital on the basis of the private control over the means of production and circulation," while informationalism "is oriented... toward the accumulation of knowledge and towards higher levels of complexity in information processing."' Cohen (n 17) 5–6 (citing Manuel Castells, *The Information Age, vol. 1: The Rise of the Network Society* (Blackwell 1996) 14–18).
[44] Cohen, *Between Truth and Power* (n 17) 7.

elevate processes of private economic ordering; it also works to *reshape government processes in their image*'.[45]

How exactly are informational capitalism and neoliberal governmentality to mirror each other? The crucial link between the two is the aim of efficiency.[46] Both informational capitalism and neoliberal governmentality strive toward efficiency. Efficiency seems like a neutral term, but under neoliberal governmentality it practically means the overall maximation of wealth through utilitarian cost-benefit calculus.[47] In any form of capitalism, the maximation of wealth is of course the outspoken aim. For the Commission, the preferred model for assessing harms and benefits is the '[c]ost benefit analysis which entails the monetization of all (or the most important) costs and benefits related to existing public intervention or all viable alternatives at hand'.[48]

By contrast, the often more explicitly stated goal of a neoliberal government is the overall maximation of welfare instead of wealth. Welfare maximization is aimed with 'cost-effectiveness analysis' which 'entails that you quantify (not monetise) the benefits that would be generated by one Euro of costs imposed on society'.[49] However, not all harms and benefits are easily measurable and accordingly, with the absence of a common 'welfare unit', human flourishing does not fit into the simple utilitarian calculus as squarely as money. Thus, welfare maximization often *de facto* ends up lopsided toward wealth maximation.[50] Welfare/wealth maximation requires the government to act like any actor in the market, commanded by utilitarian efficiency, and consequently shapes the government to an image of a company.

However, what is crucial for my thesis, is that the promotion of efficiency entails a trade-off between democracy.[51] I cannot overstate the importance of this trade-off here. It forms the backbone of everything that unfolds on the pages of this thesis. Unlike liberal governmentality that undertook the quest of decentering power to a sovereign individual,[52] neoliberalism tends to de-value the processes of a democratic state in its respective quest for the overall maximation of wealth. This is because the demands of democracy advance also non-market values and thus are not always in line with the market imperatives such as utilitarian efficiency.[53]

---

[45] ibid. Emphasis added.
[46] Britton-Purdy and others (n 38) 1790.
[47] Zachary Liscow, 'Is Efficiency Biased?' (2018) 85 The University of Chicago Law Review 1649, 1658.
[48] European Commission, 'Better Regulation Toolbox, Tool #57: Analytical methods to compare options or assess Performance' 451, <https://bit.ly/2XjYVaO> accessed 25 July 2020. See also Orla Lynskey, 'Regulating "Platform Power"' (2017) LSE Law, Society and Economy Working Papers 1/2017, 28 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2921021> accessed 25 May 2020.
[49] European Commission (n 48) 452.
[50] Britton-Purdy and others (n 38) 1797. The underlying reasoning behind the overall wealth maximation is that when the overall 'cake' grows, it can then be redistributed under the concept of fairness, that is, through taxes and income transfers, thus benefitting everyone and not just those that have gained most from the overall growth. Whether the ideal works in practice is arguably highly context-dependent. Efficient policies are not always offset for the worse-off. Liscow (n 47) 1664, 1696, 1703–1704.
[51] Britton-Purdy and others (n 38) 1813–1814, 1827–1832.
[52] Cohen (n 17) 7.
[53] David Singh Grewal and Jedediah Purdy, 'Introduction: Law and Neoliberalism' (2014) 77 Law & Contemporary Problems 1, 3–6.

Consequently, the trade-off also means shifts in power relations, channeling power away from people toward the market, *within* which the government acts.

According to Cohen: 'Transforming government in the image of markets is not an abstract exercise. It requires changes in the nature and operation of the institutions and practices that comprise government, including (…) legislatures, courts, and legal doctrines'.[54] It brings techniques that mimic market techniques and company management into government.[55] Managerial techniques tend to assert 'the superiority of efficiency, technocratic oversight, and utilitarian methods for assessing costs and benefits over fairness, openness and public-facing accountability'.[56] However, because changes in the institutions of a government must happen through law, it follows that law also must change to accommodate the new managerial techniques.[57] Therefore, on the most practical level, to implement managerial techniques the regulatory approaches are 'procedurally informal, standard-based, [and] mediated by expert professional networks'.[58] All of them are leaning toward efficiency and away from democratic accountability.

However, I am stressing that law does not only facilitate political economy but also justice. Law should neither be cynically equated with a conduit of instrumentarian political-economic ends nor entirely substituted with justice ideals.[59] Thus, fundamental rights, including the right to freedom of expression and information, can (and should) contribute to justice. More specifically, I am taking the conceptualization of fundamental rights as 'avenues of political equality' as outlined by Eleni Frantziou in her recent work on the theory of horizontality of fundamental rights in the EU.[60] Drawing upon the theories of Hannah Arendt, Jürgen Habermas and Claude Lefort, Frantziou derives the justification of fundamental rights structure from its purpose of 'creating the necessary conditions for democratic participation'.[61] A fundamental right is understood in Arendtian way, as a *capacity* to act (politically).[62] As Ingram states, fundamental rights 'are the product of past struggles and the object of present ones'.[63] She further notes that 'beyond any substantive benefits [fundamental rights framework] offers, the fundamentality of a rights system consists in its role of organising public life and in enabling people to shape and re-interpret its parameters' in an equal manner.[64] The political re-interpretation

[54] Cohen (n 17) 7.
[55] ibid.
[56] ibid 194.
[57] ibid 2, 7–8.
[58] ibid 172.
[59] Mireille Hildebrandt, *Smart Technologies and the End(s) of Law* (Edward Elgar Publishing 2015) 147.
[60] Frantziou (n 8) 175.
[61] Ibid 162.
[62] Charles Barbour, 'Between Politics and Law: Hannah Arendt and the Subject of Rights' in Marco Goldoni and Christopher McCorkindale (eds), *Hannah Arendt and the Law* (Hart Publishing 2012) 314–315.
[63] James D Ingram, 'What Is a "Right to Have Rights"? Three Images of the Politics of Human Rights' (2008) 102 American Political Science Review 401, 411.
[64] Frantziou (n 8) 163.

is aimed towards the very substance and scope of rights themselves.[65] Therefore, as a framework, fundamental rights based on political equality protect the very politicizing character of rights.[66]

To my mind, in life people have political expectations regarding their entitlements to material resources and/or status. Principled and value-thick fundamental right texts raise this anticipation and help it to transform to more a definite form as legal claims for rights. The claims result in conflicts with the entitlements of others, which has the potential to alter not only the distribution of entitlements and disentitlements, including by reformulating the meaning of the right in question. Therefore, fundamental rights, like any law, do not and cannot fully constrain political change.[67] Effectively, the focus of the analysis of rights structure now shifts to the power imbalances present in a public sphere. This is congenial to the 'non-domination principle' which asserts that instead of looking for a concrete infringement to someone's right, the assessment should interrogate power relations and potential abuse of power.[68] Yet in Frantziou's theory, fundamental rights doctrine should allow for the scrutiny of power imbalances irrespective of whether the relationship is between two private actors or between state and private actor.[69] For instance, a powerful actor may inhibit someone's capabilities to act politically within a public sphere, including the capacity to reformulate a right.[70] This type of power imbalance must be recognized as a fundamental rights issue.[71] The structure of the rights framework should be constructed to enable the vindication of rights to re-balance the power. In other words, it should foster democratic self-determination and, especially on transnational level, what Cafaggi and Pistor call 'regulatory capabilities'. They are 'institutional conditions for individuals, collectives, and entities to express their preferences, [and] choose alternative forms of regulation'.[72]

To sum the theoretical framework outlined and to apply it in the specific context of my thesis, I am stating that a legislative and administrative fundamental rights framework, as any regulation, mediates between political economy and political equality. This mediation informs the distribution of power in society. In terms of

[65] Ingram (n 63) 411–412.
[66] Frantziou (n 8) 164.
[67] Barbour (n 62) 315–319. It should be stressed that I don't seek to advance any idealistic notions of the subversive potential of rights. As Boonen points out, there are certainly limits to the politicising capabilities of fundamental rights, which have to do with the legal form and their co-evolution with industrial capitalism. See Christian Boonen, 'Limits to the Politics of Subjective Rights: Reading Marx After Lefort' (2019) 30 Law & Critique 179, 189–196.
[68] Bart van der Sloot, 'A New Approach to the Right to Privacy, or How the European Court of Human Rights Embraced the Non-domination Principle' (2018) 34 Computer Law & Security Review 539, 545–547. According to the commentators, both the ECtHR and CJEU have recently embraced non-domination principle in *Roman Zakharov v Russia* ECHR 2015–VIII 205; and Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others* (hereafter *Digital Rights Ireland*), EU:C:2014:238. In relation to the latter case, see Andrew Roberts, 'Privacy, Data Retention and Domination: Digital Rights Ireland Ltd v Minister for Communications' (2015) 78 Modern Law Review 535.
[69] Frantziou (n 8) 165–168.
[70] ibid 166. As an example from data protection, Mario Costeja González from Spain challenged Google Inc by invoking the right to data protection (among other claims). The Court of Justice of the European Union (hereafter the CJEU) sided with Mario Costeja, thus *reformulating the right to data protection* to accommodate something new that has since came to be known as 'right to be forgotten'. Case C-131/12 *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, EU:C:2014:317.
[71] Ingram (n 63) 405, 414.
[72] Fabrizio Cafaggi and Katharina Pistor, 'Regulatory Capabilities: A Normative Framework for Assessing the Distributional Effects of Regulation' (2015) 9 Regulation & Governance 95, 98.

method, deconstruction may serve to reveal how certain values and ideologies presented as foundational or neutral are actually founded on contradictions.[73] By analyzing the efficiency/democracy trade-off in fundamental rights structure, I believe I can deconstruct efficiency as a universal value in regulation and challenge the ultimate neoliberal claim that fundamental rights are compatible with efficiency.[74] More specifically, the old rule of law of liberal democratic state contains mandating principles, encapsulated in doctrine, which may cause friction for efficiency-oriented changes.[75] Friction may result in legal conflicts between players that represent different interests and diverging approaches to fundamental rights. In turn, conflicts may produce novel legal arrangements as compromises. The arrangements could then, at least in principle, end up working as more balanced legal structures under informational capitalism.

How then, in practice, am I scrutinizing the relevant law? As mentioned, the legal-dogmatic backdrop for assessing the structure of freedom of expression is provided by the ECHR and its interpretations by the ECtHR. However, it is advisable to take into account the relevant jurisprudence of the Court of Justice of the European Union (hereafter the CJEU) as the supreme interpreter of EU law. Complementary to supranational case-law, fundamental rights scholarship helps with the interpretation of doctrine. In addition, especially because this thesis moves partly on a more theoretical level (Chapters 2 and 3), I am drawing upon regulatory studies but also social studies, namely media and platform studies.

As regards the interpretation of the regulatory framework, my aim is to ferret out various and possibly conflicting interests behind regulation to observe how the final text resembles the outcomes these conflicts. I believe the interests become surfaced during the legislative processes. Ultimately, this allows me to distill insights on whether and how the regulation exhibits the efficiency/democracy trade-off. It follows from here that the source material contains policy documentation of relevant EU institutions, most notably the Commission, Council and Parliament taken together as the EU legislator, in accordance with the ordinary legislative procedure.[76] Yet I am trying to tread cautiously here: The legislative processes in the EU are meandering and not all the stages are public. I am seeking to avoid drawing definite causal lines between interests visible at preliminary stages and the final legal text. Some unpublicized EU documents have been obtained (as redacted) by NGOs and I have accessed them through their websites.

Lastly, I am laying out the structure of thesis. In Chapter 2, after first revisiting the general doctrine of (positive) fundamental right obligations, I believe it is important to briefly delve into content moderation and its potential harms to freedom of expression. The subsequent chapters 3 and 4 constitute the main part of this thesis and aim to interrogate answers to the research questions. Chapter 3 locates the efficiency/democracy trade-off in a

[73] Jack M Balkin, 'Deconstructive Practice and Legal Theory' (1987) 96 Yale Law Journal 743, 754–755.
[74] Samuel Moyn, *Not Enough: Human Rights in an Unequal World* (Harvard University Press 2018) 191–192. Even though Moyn remarks that such a claim is more of a caricature than reality, I believe it serves a purpose here to show that neoliberalism in an increasingly complex world cannot be just a plain deregulatory agenda that seeks to insulate economic activity from public law, including fundamental rights. Grewal and Jedediah Purdy (n 53) 13–14.
[75] Cohen (n 17) 270.
[76] Articles 289 and 294 of the Consolidated version of the Treaty on the Functioning of the European Union (hereafter TFEU) [2012] OJ C326/47.

more specific, but a still generalized, analytical framework. It interrogates the specific regulatory strategies under the interconnected concepts of cooperation and contest within the EU social regulation. Chapter 4 includes three case studies of recent social media regulation initiatives in the EU. The case studies do not claim to produce an exhaustive account of the 'regulatory field' on the subject matter, but to contextualize and exemplify the more general part of Chapter 3. Of course, a drawback of picked case-studies is that the results contain reservations in producing definite generalizations. Finally, Chapter 5 draws together the argumentation in chapters and delves into discussion developed on them. Thus, it re-engages both the fundamental rights doctrine and the background theory to assess how the relevant fundamental rights structure is mediating between efficiency and political equality and whether it should be restructured somehow.

2. Foundations: Freedom of Expression and Content Moderation

This chapter lays the essential foundation for the upcoming analysis in subsequent chapters. It first revisits the doctrine of fundamental rights structure regarding state's fundamental rights obligations. It then turns to freedom of expression on social media and outlines how companies govern through platform content moderation. Even though the focus is on free expression, as I mentioned in the introduction, several rights are at stake on platforms and it is advisable to point out the major repercussions to them too.

2.1 Fundamental Right Obligations

This sub-chapter outlines the doctrinal intricacies of positive obligations under the ECHR and the Charter. Fundamental rights law imposes both negative and positive legal obligations for states, even though traditionally the rights were thought to require only restraint from public incursions to private sphere.[77] The ECtHR introduced the concept of positive obligation in the European human rights jurisprudence early on in 1968.[78] According to Beijer, for the ECtHR a positive obligation means 'a requirement for states to take active measures to protect fundamental rights'.[79] It explains that 'states must "do something", "take action" or "undertake affirmative action"'.[80] The justification for the positive obligations is the effective protection of the rights in the ECHR.[81] So, whether one likes it or not, positive obligations are indispensable because, as mentioned in Chapter 1.1.2, any regulation needs monitoring and enforcement. Fundamental rights as law would have little effect if public institutions would not enforce rights as legal entitlements, i.e. if the state would always literally 'do nothing' even to prevent or sanction its own agents, as the negative obligations seem

[77] Gardbaum (n 5) 396–397.
[78] *Case "relating to certain aspects of the laws on the use of languages in education in Belgium" v Belgium (Belgian Linguistic case)* App nos 1474/62 to 2126/64 (ECtHR, 23 July 1968), para 9.
[79] Malu Beijer, *The Limits of Fundamental Rights Protection by the EU: The Scope for the Development of Positive Obligations* (Intersentia 2017) 99.
[80] ibid 41.
[81] ibid 46.

to suggest. Therefore, it may sometimes be hard to meaningfully distinguish positive and negative obligations.[82]

However, often the importance of positive obligations generally and in this thesis specifically is that they serve as a doctrinal tool for state to interrogate also situations where private action implicates the rights of others. While this forges a link to the horizontal effect of fundamental rights, i.e. the applicability of fundamental rights law in private relations,[83] the scholarship is not unanimous on the relationship between the two doctrines. It has been pointed out that the positive obligations have in practice resulted to the adoption of horizontal effect. However, Lavrysen has insisted that positive obligations and horizontal effect are 'analytically distinct'.[84] I am following the theorization of Frantziou and Alexy, who practically equate positive obligations with 'state-mediated horizontal effect' as opposed to indirect and direct horizontality.[85] The clearest distinction, and important for this thesis, is that like (classical) liberal governmentality in general, *positive obligations presuppose the strong stewardship of state*, that is, they are not imposed on private actors.[86] Congruently, the infringements on rights by private individuals or legal persons can be projected against state through the lack of preventative or remedial action.[87] This is already indicated by the term *state*-mediated horizontality. Moreover, they are often fulfilled by the actions of the legislator and the judges, while indirect and direct effect consider how rights are applicable in courts.[88] State administration has an important role in positive obligations too.

The ECtHR imposed the first positive obligations in landmark cases of *Marckx v Belgium*, *Airey v Ireland*, *Young, James and Webster v The United Kingdom*, and *X and Y v the Netherlands*.[89] These cases solidified the concept as an inherent part of the Convention system. The ECtHR has refrained from defining the positive obligations on general level. Instead, it has stressed the contextual analysis for specifying the concrete obligations.[90] This is inexorable in a sense that, while it is clear that positive steps require *some* measures to be taken by state, it is impossible to say *what* measures should be taken if the ECtHR is to take into account

[82] 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) 15; and Robert Alexy, *A Theory of Constitutional Rights* (Julian Rivers tr, OUP 2002) 359–361.
[83] Alexy (n 82) 354–355.
[84] Laurens Lavrysen, *Human Rights in a Positive State: Rethinking the Relationship between Positive and Negative Obligations under the European Convention on Human Rights* (Intersentia 2016) 79. Similarly, Gardbaum (n 5) 397.
[85] Frantziou (n 8) 44–45; and Alexy (n 82) 356–357.
[86] According to Lavrysen, the doctrine of horizontal effect is a matter to be solved on the national level '[s]ince the ECHR only imposes obligations on States and not on individuals'. Lavrysen (n 84) 13. Indeed, the Court itself has not explicitly embraced the horizontal effect doctrine but has left the extent of horizontal effect to be decided on state level. *VgT Verein Gegen Tierfabriken v Switzerland* ECHR 2001–VI 243, para 46; and *Söderman v Sweden* ECHR 2013–VI 203, para 79.
[87] Akandji-Kombe (n 82) 14–15; and Stéfanie Khoury and David White, *Corporate Human Rights Violations: Global Prospects for Legal Action* (Routledge 2017) 110.
[88] Alexy (n 82) 357.
[89] Beijer (n 79) 38–39; *Marckx v Belgium* App no 6833/74 (ECtHR, 13 June 1979); *Airey v Ireland* App no 6289/73 (ECtHR, 9 October 1979); *Young, James and Webster v The United Kingdom* App nos 7601/76 and 7806/77 (ECtHR, 13 August 1981); and *X and Y v the Netherlands* App no 8978/80 (ECtHR, 26 March 1985).
[90] Beijer (n 79) 87. Eg, *Plattform "Ärtze für das Leben" v Austria* App no 10126/82 (ECtHR, 21 June 1988), para 31.

the individualities of the case at hand. Nevertheless, legal scholarship has identified several typologies that helpfully bring flesh to its bones in the form of more crystallized obligations.[91]

Firstly, the ECtHR has recognized states having positive obligations especially in situations of horizontal relationship between private parties.[92] Therefore, following the typology of Lavrysen, positive obligations may be divided to horizontal and vertical ones.[93] According to Lavrysen, '[h]orizontal positive obligations are those positive obligations that govern relations between private persons. They are typically triangular in character, since the individual invokes them against the State to oblige State authorities to intervene in horizontal relations'.[94] Vertical positive obligations directly govern the relations between the individual and the State.[95] Akandji-Kombe calls it a 'duty of schizophrenia' as the state is required to take action to prevent or punish the violations that its own agents commit.[96] When state takes measures to interfere in horizontal relations, this may, however, infringe the rights of others. Therefore, the extent of measures is in many cases determined through *ad hoc* balancing.[97] The ECtHR refers to this balancing as 'fair balance' test and applies it generally for 'determining whether or not a positive obligation exists' in the first place.[98] Thus, the concept of fair balance is the convergence point between negative (restraint) and positive obligations (affirmative action).

Positive obligations have been further concretized by dividing them to the ones that require states to set up the appropriate 'legislative and administrative framework' and to the ones that are to be formulated on more *ad hoc* basis.[99] In case *Centro Europa 7*: 'The Court observes that in such a sensitive sector as the audiovisual media, in addition to its negative duty of non-interference the State has a positive obligation to put in place an appropriate legislative and administrative framework'.[100] Indeed, Xenos argues that setting up both the legislative and administrative frameworks for the protection against infringements by private actors are the two elements comprising 'the core content' of positive obligations.[101] Instead, *ad hoc* obligations are usually required on top of the framework and derive even more from the context of the case at hand.[102]

---

[91] See eg, Lavrysen (n 84); and Dimitris Xenos, *The Positive Obligations of the State under the European Convention of Human Rights* (Routledge 2012).
[92] Beijer (n 79) 47.
[93] Lavrysen (n 84) 57. Beijer (n 79) 58, adopts the same typology.
[94] Lavrysen (n 84) 78–79, 100.
[95] ibid 79.
[96] Akandji-Kombe (n 82) 15–16. One could ask whether vertical positive obligations are actually negative obligations and, therefore, whether the class of vertical positive obligations only creates unnecessary confusion. However, the Court has not seen them redundant as, for instance, it has required the higher authorities of the State to not only respect the Convention rights themselves, but 'those authorities must [also] prevent or remedy any breach at subordinate levels.' *Assanidze v Georgia* ECHR 2004–II 221, para 146.
[97] Lavrysen (n 84) 117–118.
[98] *Rees v the United Kingdom* App no 9532/81 (ECtHR, 17 October 1986), para 37.
[99] Lavrysen (n 84) 112.
[100] *Centro Europa 7 Srl and Di Stefano v Italy* ECHR 2012–III 339, para 134. As regards positive freedom of expression obligations, see also, *Özgür Gündem v Turkey* ECHR 2000–III 1, para 43; and *Khurshid Mustafa and Tarzibachi v Sweden* ECHR App no 23883/06 (ECtHR, 16 December 2008), paras 43, 48–50.
[101] Xenos (n 91) 107–115.
[102] The ECtHR has stipulated that '[positive] obligation may involve the adoption of *specific measures*'. *Hämäläinen v Finland* ECHR 2014–IV 369, para 63. Emphasis added. Similarly, *KU v Finland* ECHR 2008–V 125, para 49.

Regarding legislative framework, 'a positive obligation arises for the state to regulate in advance individual conduct and the operation of the activities of private parties'.[103] For instance, the ECtHR has stated that positive obligation contains 'measures [which] may include both the provision of a regulatory framework of adjudicatory and enforcement machinery protecting individuals' rights'.[104] Administrative framework contains effective implementation of the legislative framework.[105] For instance, in the case of environmental hazard, the Court has specified that the administrative framework includes the 'licensing, setting up, operation, security and supervision of the [economic] activity'.[106] Of course, the obligation requiring the adequate legislative or administrative framework is closely connected to right to effective remedy in Article 13 and right to a fair trial in Article 6(1) of the ECHR.[107] Moreover, Lavrysen considers that relying on mere *ad hoc* measures through wide discretionary powers could risk undermining rule of law. Thus, a robust and sufficiently detailed legislative and administrative framework is crucial for rule of law too.[108]

As regards the EU, before Lisbon Treaty and the Charter becoming binding, its obligations were more of negative nature. Thus, the EU was only committed not to breach fundamental rights rather than actively promoting them.[109] Now Article 51(1) of the Charter, addressing the field of application, states that '[the Union] shall therefore respect the rights, observe the principles and *promote the application thereof* in accordance with their respective powers and respecting the limits of the powers of the Union as conferred on it in the Treaties'.[110] Yet this choice of wording does not self-evidently assert positive obligations in connection to the Charter.[111] According to Kuczarawy, the existence of positive obligations for the EU is 'less obvious'.[112]

However, the positive obligation doctrine receives additional backing from the so-called homogeneity clause in Article 52(3) of the Charter.[113] According to Amalfitano, the meaning of the clause within the EU legal system is that: '(i) the ECHR represents the *minimum floor* of the level of fundamental rights protection; and (ii) the Union (therefore, above all, the political institutions in the exercise of their legislative powers) is *only* allowed to increase the level of protection provided by the ECHR'.[114] The meaning and scope of the Convention rights include also the case-law of the ECtHR.[115] Therefore, by virtue of the homogeneity clause,

[103] Xenos (n 91) 110.
[104] *Hämäläinen v Finland* ECHR 2014–IV 369, para 63.
[105] Xenos (n 91) 110.
[106] *Öneryildiz v Turkey* ECHR 2004–XII 79, para 90.
[107] Lavrysen (n 84) 117.
[108] Lavrysen (n 84) 119.
[109] Andrew Williams, 'Human Rights in the EU' in Damian Chalmers and Anthony Arnull (eds), *The Oxford Handbook of European Union Law* (Oxford University Press 2015) 252.
[110] Emphasis added.
[111] Williams (n 109) 253.
[112] Kuczerawy (n 22) 155.
[113] Chiara Amalfitano, *General principles of EU law and the Protection of Fundamental Rights* (Edward Elgar Publishing 2018) 57.
[114] ibid 58. Emphasis original.
[115] Case C-205/15 *Direcția Generală Regională a Finanțelor Publice Brașov (DGRFP) v Vasile Toma and Biroul Executorului Judecătoresc Horațiu-Vasile Cruduleci*, EU:C:2016:499, para 41.

also the doctrine of positive obligations could flow to EU system of fundamental rights. Nevertheless, without the accession of the EU to the ECHR, its effect on the Union bodies is indirect.[116]

Moreover, any recognition of positive obligations is subject to the EU competence in general since, pursuant to Article 51(2), the Charter does not in itself set up any additional tasks for the Union regarding the protection of rights or otherwise.[117] Yet the EU's Digital Single Market project undoubtedly triggers the application of Charter as well.[118] Kuczerawy sees that the requirement of effectiveness stated in Articles 4(3) and 19(1) also necessitates the existence of positive obligations for Charter rights.[119] The effectiveness requirement is of course the same as the ECtHR uses to justify positive obligations.[120] Further, in the context of Internet regulation, the CJEU has affirmed the obligation to provide effective protection and fair balance through 'sufficient safeguards'.[121] Based on these findings, Kuczerawy argues, and I agree, that there indeed exists a positive obligation to guarantee the Charter rights. Instead of explicitly referred to as a positive obligation, it manifests itself in the doctrines of effective protection and fair balance. Moreover, this obligation binds not only Member States implementing the Union law, but also the EU bodies themselves as constituting a distinct supranational legislator.[122] For instance, EU regulations don't even require any implementation to come into force. So, to summarize this sub-chapter, it has been argued that not only European states but also the EU as a legislator, are understood as the primary subjects of positive fundamental obligations, including those deriving from freedom of expression.

2.2 Hard Content Moderation and e-Commerce Directive

In this sub-chapter, I first provide a brief (and necessarily limited) account of freedom of expression and social media generally. After that, the rest of this chapter considers platform content moderation. As stated, I am dividing content moderation to 'hard' and 'soft' part. The latter part of this sub-chapter focuses on hard moderation, its legal facilitation in the e-Commerce Directive, and its implications for freedom of expression and other rights. The next sub-chapter 2.3 delves into soft moderation.

Broadly speaking, there are two theory groups for understanding the value of free expression.[123] One links speech to individual autonomy, self-determination and dignity. As Woods states, 'theories in this group concern an individual's right to choose what to believe, and therefore to have access to a choice of views and

[116] Case C-617/10, *Åklagaren v Hans Åkerberg Fransson*, EU:C:2013:105, para 44.
[117] Angela Ward, 'Remedies under the EU Charter of Fundamental Rights' in Sionaidh Douglas-Scott and Nicholas Hatzis (eds), *Research Handbook on EU Law and Human Rights* (Edward Elgar Publishing 2017) 169.
[118] See European Commission, 'Shaping the Digital Single Market' (last update 19 February 2020) <https://bit.ly/3gl6mWQ> accessed 2 June 2020.
[119] Kuczerawy (n 22) 155.
[120] Beijer (n 79) 46.
[121] Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland*, EU:C:2014:238, paras 66, 69; and Case C-362/14 *Maximillian Schrems v Data Protection Commissioner*, EU:C:2015:650, para 91.
[122] Kuczerawy (n 22) 155.
[123] Woods (n 1) 107; Mir and Bassini call it 'double dimension' of freedom of expression in European legal systems. Mir and Bassini (n 11) 78–80.

the space to experiment with and to reject ideas, as well as to express him or herself'.[124] These theories make Article 10 of the ECHR to converge with Article 8.[125] The other group sees the value in the functioning of democracy as a system.[126] Here, speech operates in the public domain where the development of society and the critique of those with power is possible.[127] Scholars have highlighted the importance of classic case *Handyside*, where the ECtHR arguably embraced both theory groups.[128] As regards social media, it enables computer-mediated social interaction for people, thus creating vast and complex sociotechnical networks that form instances of the networked public sphere.[129] They allow for a myriad of ways to form new affinity networks, the effects of which have been lauded for their enormously democratizing effects in the domains of culture and politics. Also, they have empowered people to circumvent the traditional gatekeepers of mass media, providing a chance for self-realizing participation to millions or even billions.[130]

While decentralized, networks still allow access points for power concentration.[131] The platform is a hub in a network, through which all the social interactions must flow. By controlling the hub, i.e. gatekeeping, the platform company can record all interactions and store them as data. Data forms the core of social media political economy.[132] However, people's social interactions are messy and incommensurate as they appear 'offline', and thus they must first be made commensurate to be recorded as simple numeric data. So, companies' political-economic interests require that a platform simplifies social interactions for datafication/commodification.[133] This practically means that a platform mediates social interactions through highly stylized user actions, such as a suite of 'likes' and 'shares' that equate with single clicks. These clicks can be turned to data much more easily, making them more viable for computational analytics.[134] When data is processed through computational analytics, information in the form of (unexpected) correlations is produced, with potentially high monetary value.[135]

Naturally, the number of computer-mediated clicks needs to be maximized. Companies do this through a suite of governance efforts that are meant to maximize people's engagement (i.e., clicks). These governance efforts

[124] Woods (n 1) 107.
[125] ibid 107.
[126] Mir and Bassini (n 11) 78.
[127] Woods (n 1) 107.
[128] Mir and Bassini (n 11) 78; Woods (n 1) 107. The ECtHR declared that '[t]he Court's supervisory functions oblige it to pay the utmost attention to the principles characterising a "democratic society". Freedom of expression constitutes one of *the essential foundations of such a society*, one of the basic conditions for its *progress and for the development of every man*.' *Handyside v the United Kingdom* App no 5493/72 (ECtHR, 7 December 1976), para 49. Emphasis added.
[129] Manuel Castells, 'The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance' (2008) 616 The ANNALS of the American Academy of Political and Social Science 78, 78–79.
[130] Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press 2006) 8–13.
[131] Reed and Murray (n 14) 159.
[132] Nick Couldry and Jannis Kallinikos, 'Ontology' in Jean Burgess, Alice Marwick and Thomas Poell (eds), *The SAGE Handbook of Social Media* (SAGE Publications 2017) 167,
[133] José van Dijck, Thomas Poell and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (OUP 2018) 37–40, 46–47.
[134] Couldry and Kallinikos (n 132) 171–173.
[135] Frank Pasquale, 'The Automated Public Sphere' in Ann Rudinow Sætnan, Ingrid Schneider and Nicola Green (eds), *The Politics and Policies of Big Data: Big Data, Big Brother?* (Routledge 2018) 111.

are social media content moderation. Companies' governance is possible because of power that flows from many different legal, economic and technical sources. It is not possible nor advisable to give a detailed analysis of all the power sources. However, I indicate some notable examples. For instance, network effects make the platform more valuable for everyone when it grows.[136] Thus, the ensuing locked-in effects make it harder for users to change the service because of the loss of valuable connections.[137] Legally, companies use boilerplate terms of contract law that are non-negotiable and assert a suite of legal entitlements for the companies as regards, for instance, the user data or their platform use.[138] Lastly, trade secrecy and intellectual property rights (hereafter IPR) shelter platform workings from scrutiny.[139]

The first part, hard content moderation, is disciplinary governance. Its sanctions comprise an escalating scale of disciplining actions: warnings, removals of specific content, and account suspensions and terminations.[140] From purely substantive fundamental rights perspective, hard content moderation is a controversial thing. Since the implications of user-generated content are myriad, hard moderation can both harm and benefit fundamental rights. As regards harms, the content posted on a platform by newly empowered people may, among others, violate other persons' privacy by defaming or disclosing private information,[141] constitute illegal discrimination in the form of hateful speech against racial minorities,[142] lead to harassing of women,[143] violate the rights of the child not only trough dissemination of child pornography but also through children's exposure to violent or otherwise harmful content,[144] and implicate the right to personal security through outright threats of violence.[145] Other types of illegal content are more closely related to public policy endeavors but nevertheless have also links to the enjoyment of individual rights. For instance, the protection of intellectual property has a connection with the right to property.[146]

Initially, no external community rules existed for social media platforms and the approach to the enforcement of any internal rules was hands-off.[147] Only after platforms' user-base grew in numbers and diversity, the

[136] Ingrid Schneider, 'Bringing the State Back in: Big Data-based Capitalism, Disruption, and Novel Regulatory Approaches in Europe' in Ann Rudinow Sætnan, Ingrid Schneider and Nicola Green (eds), *The Politics and Policies of Big Data: Big Data, Big Brother?* (Routledge 2018) 131.
[137] ibid 131–132.
[138] Cohen (n 17) 44–46; and Van Dijck, Poell and de Waal (n 133) 11–12.
[139] van Dijck, Poell and de Waal (n 133) 41; and Kapczynski (n 39) 1501–1503.
[140] Schwarz (n 34) 122, 131.
[141] On defamation, see eg, Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* (later *Facebook Ireland*), EU:C:2019:821, paras 12, 19. On privacy, see eg, *KU v Finland* ECHR 2008–V 125, paras 7, 48–49. Nonconsensual sharing of sexualized images or videos constitutes another prime example. Nicolas Suzor, Bryony Seignior and Jennifer Singleton, 'Non-consensual Porn and the Responsibilities of Online Intermediaries' (2017) 40 Melbourne University Law Review 1057, 1058–1061.
[142] UNCHR 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019) UN Doc A/74/486, 21–22; and David Kaye, *Speech police: The Global Struggle to Govern the Internet* (Columbia Global Reports 2019) 66–67.
[143] UNCHR 'Report of the Special Rapporteur on Violence against Women, Its Causes and Consequences on Online Violence against Women and Girls from a Human Rights Perspective' (18 June 2018) UN Doc A/HRC/38/47, 4–6.
[144] Eg, *KU v Finland* ECHR 2008–V 125, para 46.
[145] *Delfi v Estonia* ECHR 2015–II 319, para 145.
[146] Article 17(2) of the Charter; and *Anheuser-Busch Inc v Portugal* ECHR 2007–I 39, paras 63–65, 81.
[147] Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 Harvard Law Review 1598, 1631.

companies drafted internal community standards such as Twitter Rules and policies, Facebook Community Standards, and YouTube Community Guidelines for indicating prohibited and harmful behavior that would be sanctioned.[148] Through incremental steps reactionary to vocal demands from users, NGOs and legacy media, new prohibitions on different types of harmful and/or illegal content were added to the internal rule set.[149] Special digital tools, often referred to as 'flags', were introduced on platforms for the users to report content they deem illegal or offensive.[150] The reported content would be assessed by a company moderator and, where concluded to contradict the internal rules, it would be taken down. In severe cases, also user's social media account could be suspended or even permanently terminated.[151] Platform companies indirectly employ thousands of human moderators. They usually work for independent sub-contractors, often to multinationals such as Accenture. Moderators are lowly paid and some moderation centers are situated in low-cost countries.[152] Reflection on decisions is discouraged to make the process as mechanic as possible. That's why Schwarz calls this workforce 'proletarian judicial labor'.[153] Later, also civil society organizations and public agencies have been designated as 'trusted flaggers' and through their special flags, they can trigger a more streamlined review process based on the flaggers' presumed expertise in the field. This is meant to lead to an even speedier take-down process.[154]

Algorithmic moderation techniques are the latest addition to hard moderation. On the one hand, they contain 'hashing' based re-upload filters that automatically prevent uploading content that has been removed before. The technique compares the somewhat unique 'fingerprint' of the piece of content, i.e. hash, to ones included in a specific hash database.[155] On the other hand, platform companies deploy techniques based on machine-learning, in which case the algorithm flags content for review, or sometimes even outright removes it, basing on statistical inferences it has been trained to make.[156] When deployed, it can also refine its inferences by 'learning' from the content it goes through.

Reactionary content moderation based on reporting has been named the 'notice-and-take-down' model.[157] In Europe, the specific legal facilitator for the model has been claimed to lie in Articles 14 and 15 of the e-

[148] ibid 1629, 1632–1635. See eg, Facebook, 'Community Standards' Facebook Inc (2020) <www.facebook.com/communitystandards/> accessed 25 February 2020.
[149] Danielle Keats Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep' (2018) 93 Notre Dame Law Review 1035, 1037.
[150] Klonick (n 147) 1638–1639.
[151] ibid 1639–1641.
[152] Gillespie (n 31) 120–124; and Adrian Chen, 'The laborers who keep dick pics and beheadings out of your Facebook feed' *WIRED* (23 October 2014) <www.wired.com/2014/10/content-moderation/> accessed 16 July 2020.
[153] Schwarz (n 34) 128–129.
[154] Sebastian Felix Schwemer, 'Trusted Notifiers and the Privatization of Online Enforcement' (2019) 35(6) Computer Law & Security Review 1, 3.
[155] Gorwa, Binns and Katzenbach (n 37) 4.
[156] ibid 4–5. Although, the machine-learning algorithms do sometimes remove the content automatically, for instance, when 'the tool's confidence level is high enough that its "decision" indicates it will be more accurate than our human reviewers'. Monica Bickert and Brian Fishman, 'Hard Questions: What are we doing to stay ahead of terrorists?' *Facebook Newsroom* (8 November 2018) < https://bit.ly/3gyq8hm> accessed 17 July 2020.
[157] Kaye (n 142) 62–64; and Klonick (n 147) 1638–1639.

Commerce Directive, even though notice-and-take-down is not specifically mentioned.[158] Article 14 contains a conditional safe-harbor provision for an intermediary 'hosting service provider' that exempts a platform company from liability as regards illegal content, as long as the company is not aware of it. According to the CJEU, this effectively requires that the platform is 'neutral'.[159] In addition, the liability exemption is provided only where the company acts expeditiously to remove the content after becoming aware of it.[160] In turn, Article 15 prohibits Member States from imposing 'a general obligation' to actively search for illegal content from their platform so that the companies would not lose their 'neutrality'. Applicable to all kinds of illegal content irrespective of harm, the directive provisions are open-ended and provide little guidance for the appropriate means for achieving these outcomes.[161] The idea was that more specific requirements would be prescribed by Member States, but in practice the directive was mostly implemented plainly as it is.[162]

On top of the relaxed regulatory framework, hard content moderation was able to develop within the companies in a rather haphazard and piecemeal way through trial and error and sharing of 'best practices'.[163] Hard moderation has been criticized for undue limitations to freedom of expression. Scholarship and human rights organizations have pointed to the dangers of over-removal of legal speech, including warnings over 'censorship creep'[164] and 'collateral censorship'.[165] After all, freedom of expression protects also views that may 'shock, disturb and offend'.[166] Scholarship indicates that companies deploy a 'better safe than sorry' mentality in assessing takedowns,[167] with over-removal of algorithmic filters being a particularly conflictual topic.[168] At the same time, other scholars and commentators have reminded that hard moderation can serve the protection of many rights, and pointed to the persisting failures of the companies to stop the dissemination of content harmful to other's rights.[169] It has also been compellingly remarked that the perceived dichotomy between

[158] Lilian Edwards, ''With Great Power Comes Great Responsibility?': The Rise of Platform Liability' in Lilian Edwards (ed), *Law, Policy and the Internet* (Hart Publishing 2019) 272.
[159] Joined Cases C-236/08–C-238/08 *Google France and Google Inc. v Louis Vuitton Malletier and others* [2010] ECR I–2417, para 114; Case C-324/09 *L'Oréal SA and Others v eBay International AG and Others* (hereafter *L'Oréal v eBay*) [2011] ECR I–6011, paras 112–113, 116.
[160] Edwards (n 158) 269–270; and Patrick Leerssen. 'Cut Out by the Middle Man: The Free Speech Implications of Social Network Blocking and Banning in the EU' (2015) 6 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 99, 105, paras 28–30.
[161] Kuczerawy (n 22) 63.
[162] Edwards (n 158) 273.
[163] Gillespie (n 31) 6.
[164] Citron (n 149) 1049–1055.
[165] Jack M Balkin. 'Free Speech Is a Triangle' (2018) 118 Columbia Law Review 2011, 2016–2017.
[166] *Handyside v the United Kingdom* App no 5493/72 (ECtHR, 7 December 1976), para 49.
[167] Daphne Keller, 'Internet Platforms: Observations on Speech, Danger, and Money' (2018) Hoover Institution Essay, Aegis Series Paper No. 1807, 2–3 <https://bit.ly/31dLrhM> accessed 20 June 2020. For empirical research providing support for this popular claim in the context of copyright, see Jennifer M Urban, Joe Karaganis and Brianna Schofield, 'Notice and Takedown in Everyday Practice' (Version 2, 2016, updated March 2017) UC Berkeley Public Law Research Paper No 2755628, 41 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628> accessed 22 June 2020.
[168] Urban, Karaganis and Schofield (n 167) 116–120; and Daphne Keller and Paddy Leerssen, 'Facts and Where Find Them: Empirical Research on Internet Platforms and Content Moderation' in Nate Persily and Josh Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (forthcoming, CUP 2020) 9 <https://bit.ly/2EJAOfh> accessed 16 July 2020.
[169] Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (CUP 2019) 156–157.

freedom of expression and the protection of rights of others is false. This is because 'systemic discrimination and abuse have serious negative impacts on the agency and participation of people who experience them' and may thus impede the victims from exercising their freedom of expression.[170] For instance, scholarship has indicated that women may be at risk of being pushed out from platforms because of rampant harassment and threats.[171]

In Europe, many content-related debates have considered the extent of legal removals. On the level of doctrine, it is generally clear that illegal speech does exist – freedom of expression is not absolute. Thus, the common ground has sometimes been found in the requirement of fair balance.[172] Suzor summarizes that '[c]oncerns about underblocking hate speech and overblocking counter-speech, for example, can be expressed as a proportionality problem that is very familiar to human rights law'.[173] Another matter of course is the question of where the fair balance lies in individual cases. While hard moderation of companies may have benefits for rights, in my view, what is sometimes forgotten is the fact that hard moderation is an inextricable part of social media's political economy.[174] It exhibits the efficiency/democratic accountability trade-off in extreme. For instance, hard moderation algorithms, so preferred by the companies, are highly efficient in the long term[175] but at the same time highly inscrutable, even to the point of constituting so-called 'black boxes'.[176]

2.3 Soft Content Moderation: Personalization as Commodified Governance

If content moderation was to be equated with hard moderation, the most important ways to make content visible or invisible on platforms would be excluded. Thus, the most important part of social media governance would be left unaddressed too. The platform design and algorithmically executed processes organize information. Also, they largely constitute what a platform *is* and what it *offers* to a user.[177] Since a social media platform hosts content that users have created, the structure of the platform is the only thing there is. The central promise of a platform for its users is to organize the information in a legible way.[178] In information

[170] Nicolas Suzor and others, 'Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online' (2019) 11(1) Policy & Internet 84, 89.
[171] Frank Rikke Jørgensen and Lumi Zuleta, 'Private Governance of Freedom of Expression on Social Media Platforms' (2020) 41(1) Nordicom 51, 53–54, 62; and Suzor and others (n 170) 88–90.
[172] UNCHR (n 142) 5; and Dubravka Simonovic and David Kaye, 'UN experts urge States and companies to address online gender-based abuse but warn against censorship' OHCHR (8 March 2017) <https://bit.ly/39PADuv> accessed 5 March 2020.
[173] Suzor (n 169) 153.
[174] Schwarz (n 34) 134–137. As a slightly extreme but illuminating example, increasingly popular platform TikTok (owned by ByteDance) has told its moderators to discipline the poor and the 'ugly' users to woo new ones. Sam Biddle, Paulo Victor Ribeiro and Tatiana Dias, 'Invisible Censorship' *The Intercept* (16 March 2020) <https://tinyurl.com/y4eolz7o> accessed 3 August 2020.
[175] Michèle Finck, 'Automated Decision-Making and Administrative Law' (2019) Max Planck Institute for Innovation & Competition Research Paper No. 19-10, 1 <https://bit.ly/39SVM6P> accessed 20 June 2020. According to Finck, the three main benefits of algorithmic decision-making are 'efficiency, speed and correlations'.
[176] A 'black box' means 'a system whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other.' Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015) 3.
[177] Tarleton Gillespie, 'Platforms Are Not Intermediaries' (2018) 2 Georgetown Law Technology Review 198, 202.
[178] Cohen (n 17) 41–42, 75.

society, as Cohen states, '[t]he problem is not [information] scarcity but rather the need for new ways of cutting through the clutter'. Cohen calls this 'infoglut' and Pollicino 'information overload', meaning the unmanageably massive flows of information.[179] Infoglut is a fundamental condition of a social media platform. A platform must promote certain information and play down other information because everything cannot be shown at the same time. Abandoning these organizational activities would reveal the glut and render platform unviable for users.[180] But it is also inevitable that 'the power to include, exclude, and rank is the power to ensure that certain public impressions become permanent, while others remain fleeting'.[181] These processes necessarily make platforms 'the arbiters of truth', a title so vehemently rejected by Zuckerberg.[182] Platforms distribute highly relativized 'truth' in the form of content visibility.

Soft content moderation appears perhaps the clearest in different information 'feeds' coded to the user interface and operated by complex proprietary algorithms. In scholarship, these are called 'recommender systems' that include internal search fields, recommendation for network connections, and feeds on trending topics.[183] In addition to Facebook's News Feed algorithm, examples here are Twitter's 'Trends for you' and the YouTube page of 'Recommended' videos or its Autoplay video feed. To be sure, a user may influence her social media feeds through all her recorded and immediately analyzed behavior on the platform and Internet in general, for instance, by clicking the respective buttons of following, liking and so on. However, in practice it is often unclear how the user actions fuse with the other countless factors of the algorithm that conjure the information to the interface.[184] The factors, like platform design in general, are also perpetually fine-tuned by the companies.[185]

While the platform design and algorithms are easy take for granted, one should keep in mind that the means of people's communication not so much reflect the social interaction but instead they *shape* it.[186] More specifically, the mediation of information and expressions on a platform make is possible to 'nudge' users' behavior.[187] Nudging means the pre-emptive construction of situations where a person must make a choice to

[179] Cohen (n 17) 75; and Oreste Pollicino, 'Metaphors and Judicial Frame: Why Legal Imagination (Also) Matters in the Protection of Fundamental Rights in the Digital Age' in Bilyana Petkova and Tuomas Ojanen (eds), *Fundamental Rights Protection Online* (forthcoming, Edward Elgar Publishing 2020) 30 (in manuscript).
[180] Suzor (n 169) 163.
[181] Pasquale (n 176) 14. Gillespie calls this 'moderation by design'. Gillespie (n 31) 178.
[182] See eg, Tom McCarthy, 'Zuckerberg says Facebook won't be 'arbiters of truth' after Trump threat' *The Guardian* (28 May 2020) < https://bit.ly/3gnnTO6> accessed 16 July 2020.
[183] Paddy Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (SSRN 2020) 3–5 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3544009> accessed 17 July 2020.
[184] Dijck, Poell and de Waal (n 133) 41–44; and Pasquale (n 135) 115. The factors may count in thousands as in the case of Facebook News Feed. Victor Luckerson, 'Here's how Facebook's news feed actually works' *TIME* (9 July 2015) <https://time.com/collection-post/3950525/facebook-news-feed-algorithm/> accessed 19 July 2020.
[185] Dijck, Poell and de Waal (n 133) 41. For instance, the presentation of recommendations on YouTube has arguably been tweaked 'hundreds' of times in 2018 only. Leerssen (n 183) 29.
[186] Couldry and Kallinikos (n 132) 175; and Dijck, Poell and de Waal (n 133) 34–35.
[187] Daniel Susser, Beate Roessler and Helen Nissenbaum, 'Online Manipulation: Hidden Influences in a Digital World' (2019) 4(1) Georgetown Law Technology Review 1, 22–26. The term nudge has been coined by Cass R Sunstein and Richard H Thaler, *Nudge: Improving Decisions About Health, Wealth and Happiness* (revised edn, Penguin 2012), meaning 'any aspect of choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives' (7).

increase the likelihood that she chooses the action she is wanted to choose.[188] For instance, to increase the engagement on platforms, the companies can make the user choice of liking content more 'available' by placing the digitized 'like' button on the right hand corner on the users' interface rather than on the left.[189] People's choice-making and engagement levels are experimented by continuous A/B testing on platforms, naturally without people's knowledge.[190]

Platform companies do not speak of this governance as content moderation. While the purpose of the design and all the algorithmic ordering of content is to serve the companies' political-economic interests,[191] for the users, this non-disciplinary governance is wrapped in the language of personalization and actively marketed as 'features'. According to Zuboff, this marketing of personalization appeals to people's yearning for self-realization.[192] Personalization is framed as better user experience and coupled with promises to show information that user, as a unique individual, purportedly finds interesting.[193] Thus, in a way, soft content moderation is the *commodity* the platform offers.[194] For instance, Facebook Data Policy stipulates that,

> We use the information we have to deliver our Products, including to personalize features and content (including your News Feed, Instagram Feed, Instagram Stories and ads) (…) [t]o create *personalized Products that are unique and relevant to you*. [W]e use your connections, preferences, interests and activities based on the data we collect and learn from you and others.[195]

Yet soft moderation entails several threats to freedom of expression, irrespective of whether one prefers reducing its value to individual self-determination or systemic democracy. As regards individual autonomy, the coded structure of a social media platform may be designed so that it threatens rights.[196] Technologies afford different opportunities for social action and constrain others and their potential (ab)uses may constitute threats to fundamental rights.[197] For instance, while some nudges are innocuous, nudging may be used for suspect and self-serving purposes as well. Importantly, the incessant corporate surveillance and its predictive, real-time analysis lay the basis for the personalization of a user's interface. The predictive correlations are fed

[188] Karen Yeung, ''Hypernudge': Big Data as a Mode of Regulation by Design' (2017) 20 Information, Communication & Society 118, 120–121.
[189] Dijck, Poell and de Waal (n 133) 11.
[190] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power* (Profile Books 2019) 298–303. By one definition, A/B testing means 'a method for determining the relative effectiveness of messaging strategies or user interface elements by presenting different variations to different groups of users and measuring user response'. Susser, Roessler and Nissenbaum (n 187) 31 n105.
[191] Couldry and Kallinikos (n 132) 171–174.
[192] Zuboff (n 190) 30–37, 45, 256–257.
[193] Gillespie (n 177) 202–203; and Pasquale (n 176) 61.
[194] Gillespie (n 31) 13, 41–42.
[195] Facebook, 'Data Policy' *Facebook Inc* (19 April 2018) <www.facebook.com/about/privacy> accessed 24 February 2020. Emphasis added.
[196] Julie E Cohen, 'Affording Fundamental Rights: A Provocation Inspired by Mireille Hildebrandt' (2017) 4 Critical Analysis of Law 78, 84–87.
[197] Hildebrandt (n 59) 169–172; and Christoph B Graber, 'Artificial Intelligence, Affordances and Fundamental Rights' in Mireille Hildebrandt and Kieron O'Hara (eds) *Life and the Law in the Era of Data-driven Agency* (Edward Elgar Publishing 2020) 194, 196–198.

back to the user as tailored information configurations.[198] Surveillance-driven, personalized information may significantly increase the effectiveness of nudging, upgrading it to what Yeung calls a 'hypernudge'.[199] Sophisticated behavioral data analysis and platform tools afford hidden manipulative practices that target individual vulnerabilities without people's awareness.[200] For instance, advertisers can use Facebook's platform to target teenagers precisely at moments when they feel most insecure.[201] Also, its efficient advert tools allow feeding of mis/disinformation adverts precisely for those audience segments that are predicted to be prone to the messages of pseudoscience.[202] Effectively, personalization imbued with political-economic micro-targeting can be directed to work in the interest of anyone with deep enough pockets. Zuboff calls this instrumentarian power: It is effective, neoliberally blind for motives, and detrimental for individual self-determination.[203]

As regards the value of democracy as a system, Cohen sees that '[p]latform-based intermediation alters collective behavior in ways that have begun to produce large-scale societal effects'.[204] Some effects of soft moderation gather force and become visible and harmful on the collective level, threatening the democracy itself. Firstly, personalization 'intensifies in-group effects, reinforcing existing biases, inculcating resistance to facts that contradict preferred narratives, and encouraging demonization and abuse of those who hold opposite beliefs and political goals'.[205] For the feed algorithms, engagement equates with relevance, making platforms deliver whatever its respective audience segments deem engaging.[206] Pervasive personalization of information creates echo chambers, perhaps more often discussed as filter bubbles.[207] Echo chambers fuel tribalism and trigger rejection and anger toward conflicting views.[208] Thus, they reduce solidarity and can erode the basis for collective meaning-making.

Secondly, 'relevance' is especially high with negative or angry expressions, or via content otherwise located to the margins of appropriate.[209] While some are undoubtedly appalled by controversial content, many are

[198] Dijck, Poell and de Waal (n 133) 41–42; and Zuboff (n 190) 271–282.
[199] Yeung (n 188) 121–122.
[200] Susser, Roessler and Nissenbaum (n 187) 43; Yeung (n 188) 124; and Zuboff (n 190) 307–309.
[201] Zuboff (n 190) 305–307.
[202] Nathalie Maréchal, Rebecca MacKinnon and Jessica Dheere, 'Getting to the Source of Infodemics: It's the Business Model' *Ranking Digital Rights* (26 May 2020) 17–18.
[203] Zuboff (n 190) 352, 376–382. However, the exact severity of the threat to autonomy, most notably argued by Zuboff, is not clear. In other words, the effectiveness of behavioral micro-targeting for individual decision-making may have been exaggerated by some proponents and critics alike. This of course does not mean that manipulative practices would not be a threat and even more in the future. See Kapczynski (n 39) 1472–1474, for a summary of these doubts.
[204] Cohen (n 17) 86.
[205] ibid 87.
[206] Jack Andersen and Sille Obelitz Søe, 'Communicative Actions We Live by: The Problem with Fact-checking, Tagging or Flagging Fake News – The Case of Facebook' (2020) 35 European Journal of Communication 126, 134.
[207] Eli Pariser, *The Filter Bubble: What the Internet is Hiding from You* (Penguin 2012) 9. He states that personalization algorithms 'create a unique universe of information for each of us (…) you're the only person in your bubble'. Cohen sees the filter bubble a somewhat misleading metaphor since people not so much filter into individual bubbles but rather into collective tribes that correspond with their respective inclinations. Julie E Cohen, 'Tailoring Election Regulation: The Platform Is the Frame' (2020) 4 Georgetown Law Technology Review 641, 646–647.
[208] Cohen (n 207); and Pasquale (n 135) 120–121.
[209] James Grimmelmann, 'The Platform Is the Message' (2018) 2 Georgetown Law Technology Review 217, 227; and Nathalie Maréchal and Ellery Roberts Biddle, 'It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge' *Ranking Digital Rights* (17 March 2020) 22.

captivated by it. Thus, social media platforms need also content that leans towards extremes.[210] But amplified visibility also greatly amplifies the harms that content may cause to others. In addition, as on the individual level, platform's affordances of 'diffusion, polarization, and relativization are easily manipulated and weaponized'.[211] As Yeung argues, hypernudging is not 'hyper' only because it's more effective on the individual level, but because it allows real-time nudging of huge masses. The result is the threat of manipulation *at scale*.[212] Early on, public relations industry was apt in reining in the instrumentarian platform power, creating a rent-seeking 'meme-industry' that produces social media virality in the interest of agencies' clientele.[213] But nowadays more malevolent actors have become to master the affordances for collective manipulation and harm amplification too (with or without paying). These include not only suspect extremist organizations of all sorts, but also foreign governments.[214] For instance, China has been revealed to run social media disinformation campaigns in Europe during the COVID-19 pandemic.[215] Some disinformation campaigns may seek to manipulate democratic elections.[216]

Thirdly, polarization coupled with the amplification of scandalous, engaging content makes it easy to inflame collective outrage, leading to targeted hate campaigns against minorities, politicians and ordinary citizens.[217] Collective outrage combined with popular (and sometimes deeply needed) social media activism has also given rise to neoliberal 'cancel culture' which has proved highly lucrative for public relations industry offering 'PR crisis management'.[218] Yet it can also shun people from participation.[219] Lastly, sophisticated advert tools conveniently allow discriminatory micro-targeting, recasting the poor and minorities as 'low-value individuals', thus providing even more ways to entrench the existing forms of disentitlement in society.[220]

This reveals how schizophrenic (and perhaps fruitless) an exercise preventing the dissemination of illegal content is for platform companies. They are striving to curtail speech whose unprecedently effective dissemination is largely caused by the platforms' own processes.[221] Moreover, the company attempts to address dangerous content have raised their respective controversies. For instance, freedom of expression may be implicated when companies tweak the algorithmic weights to render valuable (but less-engaging) voices

[210] Grimmelmann (n 209) 227–228.
[211] Cohen (n 17) 87.
[212] Yeung (n 188) 122–123.
[213] Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (Public Affairs 2013) 154–158.
[214] Cohen (n 17) 88–89.
[215] Jennifer Ranking, 'EU says China behind 'huge wave' of Covid-19 disinformation' *The Guardian* (10 June 2020) <https://bit.ly/3hXZbUQ> accessed 2 August 2020.
[216] Tom Dobber, Ronan Ó Fathaigh and Frederik J Zuiderveen Borgesius, 'The Regulation of Online Political Micro-targeting in Europe' (2019) 8(4) Internet Policy Review 1, 2–5.
[217] Cohen (n 17) 88–89.
[218] Helen Lewis, 'How capitalism drives cancel culture' *The Atlantic* (14 July 2020) <https://bit.ly/39O2Z8h> accessed 19 July 2020.
[219] See eg, Elliot Ackerman and others, 'A letter on justice and open debate' *Harper's Magazine* (7 July 2020) <https://harpers.org/a-letter-on-justice-and-open-debate/> accessed 19 July 2020.
[220] Karen Yeung, 'Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism' (2018) 8 International Data Privacy Law 258, 264–266.
[221] Grimmelmann (n 209) 217.

practically invisible. For instance, social media platform's internal search algorithm may be re-calibrated so that some search queries do not generate any results even though the content itself is still on the platform as before.[222] Also, Facebook used to run a fact-checking project where content designated 'fake' received 80% of reduction in visibility in the News Feed algorithm.[223] Arguably, from similar visibility downranking suffer also otherwise 'inauthentic' or 'low-quality' content, whatever these company-stamped categories entail.[224]

The amplification of harmful content also brings up a peculiar challenge of 'probabilistic free speech'.[225] The knotty issue of probabilistic free speech is to determine whether, say, a 80% reduction in visibility on a platform constitutes a proportionate limitation to free expression, or what are the practical results of an 80% reduction in the first place.[226] Is it already invisible? Intuitively, one could see that there are legitimate interests to scale down the virality of harmful expressions, and a person's freedom of expression cannot include *a priori* entitlement to better visibility than other expressions.[227] But some benefit from this governance while others must bear their adverse effects, and the determinations are made by company policy experts who impose these probabilistic consequences.[228]

An interesting link between hard moderation and soft moderation is forged when platform companies discipline users for strategic 'gaming of the system'. While some of these efforts rightly target bots, disinformation and hateful propaganda, also cultural workers, which have become highly dependent on social media visibility, have been sanctioned for trying to figure out the algorithmic workings to ensure visibility for their expressions.[229] Those users face warnings and ultimately content/account removals. In reality, research has indicated that the line with 'genuine' and 'ingenuine' content often far from clear.[230] The accusations of system-gaming are coupled with its even more villainous counterpart 'system-abuse'.[231] As the companies strategically insist that algorithmic visibility is somehow meritocratically earned, the gaming of users by the companies themselves is left unaddressed.[232] In reality, a platform has been designed for gaming, it just wasn't expected to be used in the interest of others than the platform companies and their business partners.

---

[222] Gillespie (n 31) 182–190.
[223] Mike Ananny, 'Making up Political People: How Social Media Create the Ideals, Definitions, and Probabilities of Political Speech' (2020) 4 Georgetown Law Technology Review 351, 362–363.
[224] ibid 359–364. On the political nature of the categories, see also Andersen and Obelitz Søe (n 206) 127.
[225] Mike Ananny, 'Probably speech, maybe free: toward a probabilistic understanding of online expression and platform governance' *Knight First Amendment Institute* (21 August 2019) <https://bit.ly/31ckHhW> accessed 19 July 2020.
[226] Ananny (n 223) 363–365.
[227] Renee Diresta, 'Free speech is not the same as free reach' *WIRED* (20 October 2018) <www.wired.com/story/free-speech-is-not-the-same-as-free-reach/> accessed 19 July 2020.
[228] Ananny (n 225).
[229] Caitlin Petre, Brooke Erin Duffy and Emily Hund, '"Gaming the System": Platform Paternalism and the Politics of Algorithmic Visibility' (2019)(October–December) Social Media + Society 1, 1–3.
[230] ibid 5–7.
[231] Jessica Romero, 'Taking legal action against those who abuse our services' *Facebook Newsroom* (18 June 2020) <https://about.fb.com/news/2020/06/automation-software-lawsuits/> accessed 19 July 2020.
[232] Petre, Duffy and Hund (n 229) 8–9. As an example of meritocracy claims, when Twitter banned all political advertising in late 2019, the founder and CEO Jack Dorsey declared: 'We believe political message reach should be earned, not bought.' As cited in Julia Carrie Wong, 'Twitter to ban all political advertising, raising pressure on Facebook' *The Guardian* (20 October 2019) <https://bit.ly/39Qgnc2> accessed 19 July 2020.

In sum, the social media platforms have produced a structure for the networked extensions of European public spheres that are impaired by structural dysfunctionalities and infested with rent-seeking and threats to several rights. At the same time, algorithmic intermediation of information has proved extremely efficient and lucrative for the companies.[233] Yet it's usually very hard for an individual to observe the relatively poorly understood and intangible harms of soft moderation and, thus show an infringement to mount challenge against these techniques. For instance, manipulative online practices are extremely hard to discern.[234] Correspondingly, Pariser notes that the bias of filter bubbles is almost impossible to see 'within the bubble'.[235] As regards soft moderation specifically, so far the EU has introduced only some bits of high-level and non-binding regulation.[236]

3. General Framework of Cooperation and Contest

3.1 Decentring Positive Obligations

Thus far, we have asserted that EU and Member States have positive freedom of expression obligations to install an adequate legislative and administrative framework to protect people against state but also against private actors. We also explored the wealth of harm for free expression (and other rights) present on social media platforms. Harms flows interconnectedly from user action that platforms have empowered and from the hard and soft moderation mechanisms of platforms. These are the issues that the framework should account for. In this chapter, I begin exploring how social media and freedom of expression are being regulated in the EU and how the regulation metes out power between different actors.

I start from the argument that effective democratic control over users' and platform companies' behavior is diminished. Regulatory studies have theorized models for effective control in such situations. As for the institutional forms of regulation, Julia Black's theory on 'decentred regulation' is one of the most influential and, arguably, one of the most relevant for cyberspace regulation.[237] According to Black, decentring 'is used to express the observation that governments do not, and the proposition that they should not, have a monopoly on regulation'.[238] Positively, other social actors than the state, ranging from industries to international organizations, have come to have the power to regulate as well. No single actor has enough power to achieve its respective aims alone by imposing regulation unilaterally. Decentred understanding of regulation foresees complex interactions and interdependencies between the government and other actors, and thus, also the border between private and public collapses.[239] However, from the decentring of regulation it also follows that there

[233] Margi Murphy and others, 'Tech results: Amazon, Apple and Facebook earn record profits during pandemic' *The Telegraph* (20 July 2020) <https://bit.ly/2DrFt4Y> accessed 2 August 2020.
[234] Susser, Roessler and Nissenbaum (n 187) 27, 41.
[235] Pariser (n 207) 10.
[236] See eg, EU Code of Practice on Disinformation (2018) <https://bit.ly/33hRTHt> accessed 19 July 2020.
[237] Reed and Murray (n 14) 143–144.
[238] Julia Black, 'Decentring Regulation: Understanding The Role of Regulation and Self-Regulation in a 'Post-regulatory' World' (2001) 54 Current Legal Problems 103, 103.
[239] ibid 106–111.

is an increasing pressure to decentre fundamental rights governance too, as Laidlaw has shown.[240] This concerns especially the traditionally state-focused positive obligations. In this sub-chapter I first briefly explore the challenges behind decentring and then define how this decentring leads to power struggles.

3.1.1 Challenges for Regulating Social Media

Decentring underlines the real practical constrains that a public regulator has come to have when trying to impose its will on people's behavior on social media. It would be intellectually dishonest to overlook these arguments as mere masks of neoliberalism. Thus, as stated, EU's ability to unilaterally protect rights by regulating social media is diminished. More specifically, I recognize three impediments which I call the realities of social media. While these impediments are real, I am not arguing that they appeared naturally, but they are manifestations of the political-economic advancement of informational capitalism. The three challenges are the challenge of scale, challenge of transnationalism, and challenge of complexity.

The first challenge refers to the massive scale on which the platforms operate. For instance, at the end of last year, Facebook had 2,5 billion monthly active users[241] and 100 billion pieces of content are posted on the platform each day.[242] Similarly, YouTube has over 2 billion monthly users and one million hours of video is watched each day on the platform.[243] Thus, any regulatory solution to the threats of illegal content, or to the threat of content moderation itself, should take into account the enormous scale of platforms.[244] However, the argument of scale has also been deployed purposefully to imply that platforms are beyond effective regulation, at least without impossibly high costs.[245]

Secondly, decentring theory is specifically meant to answer the effects of globalization.[246] The challenge of transnationalism may be said to follow directly from the former challenge, as the platforms' data analysis business necessitates the economies of scale.[247] Thus, the platform scaling cannot be confided to any single nation state.[248] While the biggest social media platforms were all founded in the US, they nowadays operate globally.[249] This means that attempts to regulate often extend only to a state's 'respective part of cyberspace', or slightly further to online activities which have a direct link to that state.[250] Since platform companies operate globally but lack establishment in most countries, many smaller states lack the sufficient leverage to control

[240] Laidlaw (n 15) 232–233; 236–238. Her analysis is in global context, which does not diminish its relevance here.
[241] Facebook Inc, 'Facebook Reports Fourth Quarter and Full Year 2019 Results' (29 January 2020) <https://bit.ly/3i5FCdr> accessed 14 April 2020.
[242] Patrick Wintour, 'Mark Zuckerberg: Facebook must accept some state regulation' *The Guardian* (15 February 2020) <https://bit.ly/33pQ3nI> accessed 14 April 2020.
[243] YouTube, 'YouTube for the Press' <www.youtube.com/about/press/> accessed 14 April 2020.
[244] Paul Ohm, 'Regulating at Scale' (2018) 2 Georgetown Law Technology Review 546, 546–547; and Suzor (n 169) 159–161.
[245] Cohen (n 207) 660–661.
[246] Black (n 238) 104.
[247] Zuboff (n 190) 200–202.
[248] ibid 82–83.
[249] Gillespie (n 31) 35–36.
[250] Reed and Murray call this 'the state's extended community in cyberspace'. Reed and Murray (n 14) 21.

the companies.[251] Only regulators that exercise control over the most valuable markets are usually a worthy match for the companies.[252] The EU is arguably one of them. Yet transnationalism has been manifested before the CJEU as thorny problems of global jurisdiction.[253] Jurisdictional issues are also present within the Union.[254]

Lastly, 'a system needs to be understood to be governed'.[255] The challenge first refers to the notion of technical complexity. The workings of platform algorithms and design code that curate, amplify, report, or filter content, are inherently information-dense.[256] Yet ample scholarship is pointing how the opaqueness of platform mechanisms, resulting from self-serving industry secrecy and obfuscation, is hampering regulatory efforts too.[257] From early on, opaqueness was intentionally created and nurtured to shelter the emergent data-based business model from regulatory scrutiny.[258] This is illustrated through the notion of information asymmetry, which exists not only between the citizens and companies, but informs also the relationship between regulators and companies. In one account, information asymmetry is summarized as follows:

> At the moment, determining exactly how to change these systems requires insight that only the platforms possess. Very little is publicly known about how these algorithmic systems work, yet the platforms know more about us each day, as they track our every move online and off.[259]

Decentred regulation paradigm recognizes that many social issues are so complex that no single actor has the informational capacity for successful unilateral regulation. In other words, knowledge is fragmented.[260] Finck calls to acknowledge that 'policy-makers frequently simply do not dispose of the required skillset to engage with these [social system] phenomena'.[261] Arguably, this does not only mean information asymmetry between the regulator and regulated, but underlines that not even the industry can have all the knowledge of complex social issues. Moreover, each actor understands the phenomenon through their respective lens, and thus also each actor's information on the subject matter is constructed differently.[262] But to my mind, the notion of information asymmetry should not be downplayed either. It is still hovering above regulatory attempts.

---

[251] Ohm (n 244) 549–550. For instance, already in 2014 Google closed its Google News service in Spain in response to state's intended 'link tax', which was meant to protect the country's local media. Edwards (n 158) 260.
[252] Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation' (2020) 43 Fordham Internal Law Journal 939, 952.
[253] See Case C-18/18 *Facebook Ireland*, EU:C:2019:821; and Case C-507/17 *Google LLC, successor in law to Google Inc v Commission nationale de l'informatique et des libertés (CNIL)*, EU:C:2019:772.
[254] See eg, Council of Europe, 'Liability and Jurisdictional Issues in Online Defamation Cases' (September 2019) DGI (2019) 04.
[255] Mike Ananny and Kate Crawford, 'Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973, 982.
[256] Pasquale (n 176) 61.
[257] See eg, Lynskey (n 48) 21–24; Zuboff (n 190) 88–89; Suzor (n 169) 165; and Cohen (n 17) 174–175.
[258] Zuboff (n 190) 88–92.
[259] Maréchal and Roberts Biddle (n 209) 11.
[260] Black (n 238) 107.
[261] Michèle Finck, 'Digital Co-regulation: Designing a Supranational Legal Framework for the Platform Economy' (2017) LSE Legal Studies Working Paper No. 15/2017, 19 <https://bit.ly/3fpMSPk> accessed 23 March 2020.
[262] Black (n 238) 107.

3.1.2 Cooperation and Contest

We have already explored the specific challenges that pressure the EU to adopt new decentred model for effective control over people's and social media companies' actions. However, to analyze the power implications of the ensuing regulatory strategies, I need an analytical framework with higher granularity than just the efficiency/democracy trade-off outlined in the introduction. Again, I am inspired by Cohen who takes the idea of sovereignty in exception and adjusts it to the new environment of networked and intermediated information. She states that in the information age '[s]overeignty consists in the power to say what information will flow and what will not'.[263] This power largely contains the power to dictate the scope and meaning of free expression and information: What information is blocked; what information constitutes a legal expression; what information is afforded visibility through algorithmic amplification; what information is disfavored?

As a regulator's (meaning the EU/Member States) ability to impose regulation unilaterally is diminished, rivalries sprout between the regulator and the private Internet companies to claim the sovereignty over information flows. This results in the interplay of cooperation and contest.[264] According to Cohen, regulatory hybrids are the result of this interplay. They exhibit the power of both the state and the companies.[265] Because of its cost-effectiveness, the new hybrid regulation is also in better harmony with neoliberal governmentality. Indeed, in accordance with neoliberal governmentality, decentred regulation promotes regulatory strategies that are hybrid (combining governmental and non-governmental actors), multi-faceted (using a number of different strategies simultaneously or sequentially), and indirect.[266]

However, democratic control becomes residual in cooperation and contest, which is why Cohen notes that the information sovereignty is based on exception.[267] In the sovereignty interplay, the regulator is animated by efficiency, effectively acting as a market-actor. It pursues policy aims with the most efficient means, all the while de-valuing democratic accountability. Whereas in Europe the concerns over the state of exception have lately focused on Hungary and Poland,[268] the state of exception may be worked into society more gradually and insidiously too.[269] This is squarely captured in Balkin's comparison of constitutional rot and constitutional crisis. Constitutional rot means 'a degradation of constitutional norms that may operate over long periods of time. (…) Constitutional rot is a process of decay in the features of our system of government that maintain it as a healthy democratic republic. As constitutional rot occurs, our system becomes simultaneously less

[263] Cohen (n 17) 109.
[264] ibid. For Cohen, the rivalry is between three actors, including also the intellectual property owners. However, for my purposes in this thesis the interplay of the state and platform companies is more central.
[265] ibid 121.
[266] Black (n 238) 111.
[267] Cohen (n 17) 108–109. She is here inspired by the German political theorist Carl Schmitt and Italian theorist Georgio Agamben.
[268] See eg, Case C-286/12 *European Commission v Hungary*, EU:C:2012:687; and Case C-619/18 *European Commission v Poland*, EU:C:2019:531. For a condensed summary on the EU rule of law endeavors, see European Parliamentary Research Service, 'Protecting the Rule of Law in the EU: Existing Mechanisms and Possible Improvements' (November 2019) PE 642.280.
[269] Cohen (n 17) 108. She states that 'some actions are lawlike in form but not in substance; they manifest the bare force of law stripped of the features that give the rule of law legitimacy.' (109).

democratic and less republican'.[270] In the case of free expression, it would mean a framework that relinquishes its ability to nurture the right's politicizing character.

I am taking the dialectic of cooperation and contest and applying it to the EU regulation on social media. While vying for sovereignty over information, the choice between cooperation and contest for both the companies and the EU depends on which one affords more utilitarian efficiency in a specific situation. As cooperation and contest are only different sides of the same efficiency coin, democratic accountability notions are secondary for both. However, no regulation is only a means for instrumentarian ends but affords at least some access points for political equality as well. Therefore, contest denotes another dialectic too, namely the challenge between the market interest of efficiency and political equality. In the next two sub-chapters, we set out the elements of cooperation in legislative and administrative framework. It is argued that in cooperation, the respective aims of the EU and the platform companies are thought to be aligned, making it advisable to work together to maximize efficiency. The latter two sub-chapters delve into the other half of contest.

3.2 Legislative Framework: Cooperation in Norm-setting

Cooperation is presided by the assumption that both the European regulators and platform companies share the same goal for the protection of free expression and other rights. It nurtures the idea that while the companies may have neglected looming threats, their intentions were naïve but well-meaning and 'if only people knew how to behave' we would have had an ideal public sphere. Yet the companies now genuinely want to take the responsibility for these unexpected adverse turns, at least, with some 'inducements'. It also presupposes a consensus on the meaning of freedom of expression, i.e. that regulators and everyone else see the harms and their prevention/remedy the same way, and thus the aims of the regulation are simply 'just'.

So, in scholarship, it has been remarked that already under the e-Commerce Directive the European governments tended to delegate certain important public policy tasks, mainly the protection of IPR and counter-terrorism, to platform companies. Companies have enforced these policies through hard moderation; content removals and account suspensions.[271] This has led to adverse incursions to free expression.[272] Scholars have indicated that the reason for public authorities to adopt this strategy has been the possibility to evade constitutional scrutiny.[273] Under this rationale, which Keller calls 'laundering state action', content take-downs carried out by *private companies* would not, presuming the relative weakness of horizontal effect, trigger the state responsibility in courts for removing lawful speech.[274] While it has been compellingly argued that at least

[270] Jack M Balkin, 'Constitutional Crisis and Constitutional Rot' in Mark A Graber, Sanford Levinson and Mark Tushnet (eds), *Constitutional Democracy in Crisis?* (OUP 2018) 105.
[271] Kuczerawy (n 22) 31–35.
[272] Frank Rikke Jørgensen and Anja Møller Pedersen, 'Online Service Providers as Human Rights Arbiters' in Mariarosaria Taddeo and Luciano Floridi (eds), *The Responsibilities of Online Service Providers* (Springer 2017) 180.
[273] Suzor (n 169) 152–153; Daphne Keller, 'Who Do You Sue?: State and Platform Hybrid Power over Online Speech' (2019) Hoover Institution Essay, Aegis Series Paper No. 1902, 3, 6 <https://hvr.co/39PBzz1> accessed 20 March 2020.
[274] Keller (n 273) 3.

some forms of laundered state action could nevertheless be directly attributed to the state,[275] so far the strategy has at least to some extent worked. For instance, the CJEU has struck down only the very broadest monitoring injunctions imposed on the companies by Member State courts imposed under the e-Commerce Directive.[276]

Within regulatory studies, this strategy can be theorized under 'the RIT model' where a regulator relies on 'regulatory intermediaries', such as private actors, to carry out regulatory tasks.[277] This requires that a regulator first delegates its power to an intermediary. With this strategy, the regulator can advance its interests effectively and efficiently. In turn, often the motivator for the intermediary here is the attainment of power.[278] In EU platform regulation, platform companies are these intermediaries.[279] However, the highly clinical term 'RIT model' hides the strategy's neoliberal underpinnings.[280] Thus, while I rely on the regulatory intermediary model to make sense of regulatory cooperation, I call it 'regulatory outsourcing' instead. For the regulator, outsourcing to private intermediaries promises not only a possibility to advance vital public interests without constitutional constraints. It affords significant increase in efficiency as well. In this respect, providing adequate protection for rights is no different endeavor than, for instance, advancing public security. Setting up the adequate regulatory and administrative framework for the protection of users' rights is expensive and time-consuming. Indeed, the resource-intensiveness is thought to be inherent in positive fundamental rights obligations.[281] Thus, while the EU is increasingly trying to incorporate fundamental rights safeguards to account for the older criticism, the responsibility for safeguards can be outsourced as well as another task.

According to Cohen, outsourcing presupposes that from the perspective of the outsourcing entity, the function outsourced is seen as cost-intensive, peripheral, and out of the entity's competence. In addition, 'outsourcing will be efficient only if the savings in production costs outweigh the increase in communication and coordination costs'.[282] These communication and coordination costs comprise 'transaction costs'.[283] In business management, transaction costs are calculated to determine whether an activity is on net more efficient

[275] Molly K Land, 'Regulating Private Harms Online: Content Regulation under Human Rights Law' in Rikke Frank Jørgensen (ed), *Human Rights in the Age of Platforms* (MIT Press 2019) 299–300. See also, *Sychev v Ukraine* App no 4773/02 (ECtHR, 11 October 2005), para 54.
[276] See eg, Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* (hereafter *SABAM v Netlog*), EU:C:2012:85, paras 50–52.
[277] While traditionally the presumption has been that a regulator (R) regulates people and companies as a 'target' (T), now the model includes also an intermediary (I). This results in the regulatory formula R → I → T, i.e. the RIT model. Kenneth W Abbott, David Levi-faur and Duncan Snidal, 'Theorizing Regulatory Intermediaries: The RIT Model' (2017) 610 The ANNALS of the American Academy of Political and Social Science 14, 14–15, 18.
[278] ibid, 19–20.
[279] Christoph Busch, 'Self-regulation and Regulatory Intermediation in the Platform Economy' in Marta Cantero Gamito and Hans-W Micklitz (eds), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes* (Edward Elgar Publishing 2020) 120–125.
[280] Although, the model does incorporate power considerations and Abbott and others concede that it's amenable for constructivist approaches. Abbott, Levi-faur and Snidal (n 277) 17, 28–31.
[281] Beijer (n 79) 81, 309.
[282] Cohen (n 17) 154.
[283] According to Britton-Purdy and others, 'transaction costs are costs of market exchange: of locating parties; negotiating deals; overcoming strategic bargaining problems; and, in some cases, supporting the institutions required to enable transactions'. Britton-Purdy and others (n 38) 1799.

to keep under the direct command hierarchy within the company, or to contract out to market.[284] The new technologies for information processing and communication promise significant savings in coordination costs, enabling a business organization where everything is outsourced and entirely based on networks.[285] As regards fundamental rights, their benefits don't easily lend themselves to quantified measuring.[286] Yet the costs of preventative/remedying mechanisms certainly do. Therefore, complex fundamental rights aspirations are in danger to be run over in a cost-benefit analysis, seen mainly as expenses.[287] Thus, it is tempting to conceive positive obligations as a prime target for regulatory outsourcing. There are two outsourcing strategies that can be worked into the legislative framework of social media: 1) the concept of 'voluntariness' and 2) binding norms that are principled and highly opentextured instead of precise rules.

3.2.1 Voluntariness Conundrum, Co-regulation, and Multistakeholderism

The first outsourcing method involves platform companies already at the stage of norm-setting. It relies on the strategic use of the 'voluntariness' concept.[288] Articles 290 and 291 of the TFEU set strict conditions for the delegation of rulemaking powers to the Commission. Very early on, the CJEU also imposed scrupulous requirements for the devolvement of the Commission powers further to private actors, including the prohibition to delegate rulemaking power or power that the Commission does not enjoy itself. The reasoning based on the adverse effects that power delegation has for democratic accountability and for the oversighting function of the CJEU itself.[289] In practice, often the devolvement of power to the Commission or private actors is not prescribed in EU legislation. Thus, in many technically complex sectors that are informed by high regulatory costs, the Commission has more recently pursued to facilitate 'voluntary' private regulation. It relies on insistence that rulemaking by private companies is not an unlawful delegation of power because compliance to them is voluntary.[290] In practice, this private rulemaking denotes different self/coregulatory codes of conduct. The norm-setters, in our context the platform companies, are regulatory intermediaries which largely come to draft the non-binding norms, often called 'standards',[291] on behalf of democratic legislatures.

---

[284] ibid 1799.

[285] Manuel Castells, *The Rise of the Network Society* (2nd edn, Wiley-Blackwell 2010) 176, 186–187. Castells calls this 'network enterprise' and its extreme form 'hollow corporation'.

[286] Galit A Sarfaty, 'Measuring Corporate Accountability Through Global Indicators' in Sally Engle Merry, Kevin E Davis and Benedict Kingsbury (eds), *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (CUP 2015) 122.

[287] Laidlaw (n 15) 251.

[288] Coche (n 3) 3.

[289] Rob van Gestel and Peter van Lochem, 'Private Standards as a Replacement for Public Lawmaking?' in Marta Cantero Gamito and Hans-W Micklitz (eds), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes* (Edward Elgar Publishing 2020) 32–33. Case C-9/56 *Meroni & Co Industrie Metallurgiche SpA v High Authority of the ECSC* [1958] ECR 133; Case C-10/56 *Meroni & Co Industrie Metallurgiche SpA v High Authority of the ECSC* [1958] ECR 157. More recently, however, the CJEU has ruled that the limited possibilities for delegation in Articles 290 and 291 of the TFEU do not exclude other forms of delegation, provided of course that delegation of power is clearly prescribed and delineated in an EU legislative act (34). See Case C-270/12 *United Kingdom of Great Britain and Northern Ireland v European Parliament and Council of the European Union* (hereafter *ESMA*), EU:C:2014:18, paras 79–86.

[290] van Gestel and van Lochem (n 289) 25, 31, 51.

[291] Abbott, Levi-faur and Snidal (n 277) 17.

However, scholarship has insightfully pointed out that 'voluntariness' is far from genuine if self-regulation is paired with implicit threats on binding regulation in the future. In fact, the threat on binding regulation in the future is a constitutive element of *co-regulation*.[292] Neither is the voluntary compliance plausible if a code of conduct is coupled with a strong presumption of conformity by public regulator, making the code *de facto* followed standard.[293] While the Commission advances a rather singular difference of 'pure self-regulation' and other (impure) self-regulation, where the impure version 'at the EU level (…) generally involves the Commission in instigating or facilitating the drawing up of the voluntary agreement'.[294] In any event, both self- and co-regulation shift the primary responsibility for regulation to private actors.[295]

Unfortunately, the relationship between self/co-regulation and fundamental rights has long been strained.[296] Traditionally, the principle from the early 2000s has been that these regulatory strategies are inapplicable where fundamental rights are at stake.[297] Importantly, this point was not reiterated in the revised Interinstitutional agreement on 'better' lawmaking.[298] As Marsden notes, the core question here is that 'who will decide what is so unimportant that it can be decided by co-regulation'.[299] He sees the introduction of co-regulation into fundamental rights issues resulting from the increase of commercially applied human rights law in various forms.[300] Self/co-regulation enjoy particular allure in media and communications sector.[301] The core problem of outsourcing fundamental rights norm-setting to social media companies is that co-regulation allows an access point for social media companies to define the aims of protection, i.e. which harms are recognized, prevented and redressed as fundamental rights harm.

So, 'voluntary' self/co-regulation sidelines the democratic involvement in norm-setting somewhat entirely. However, arguably this could be mended by involving all the stakeholders at the stage of drafting. It means industry but also NGOs, academia, and perhaps even ordinary citizens among others.[302] This is often meant by a rather uncrystallized term of 'multistakeholderism'. Raymond and DeNardis define multistakeholderism 'as two or more classes of actors engaged in a common governance enterprise concerning issues they regard as

[292] Chris Marsden, 'Prosumer Law and Network Platform Regulation: The Long View Towards Creating Offdata' (2018) 2 Georgetown Law Technology Review 376, 378, 394; and Citron (n 149) 1045–1049.
[293] van Gestel and van Lochem (n 289) 25, 51.
[294] European Commission (n 27) 109.
[295] Maja Cappello (ed), *Self- and Co-regulation in the New AVMSD* (IRIS Special, European Audiovisual Observatory 2019) 20.
[296] ibid 16–17.
[297] Interinstitutional agreement on better law-making [2003] OJ C321/1, para 17 affirmed that '[t]he Commission will ensure that any use of co-regulation or self-regulation is always consistent with Community law and that it meets the criteria of transparency (in particular the publicising of agreements) and representativeness of the parties involved. (…) *These mechanisms will not be applicable where fundamental rights or important political options are at stake*'. Emphasis added.
[298] Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European Commission on Better Law-Making [2016] OJ L123/1.
[299] Christopher T Marsden, *Internet Co-regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (CUP 2011) 58.
[300] ibid. Most notable here is the global trend of corporate social responsibility (CSR) that deploys non-binding norms for fundamental rights protection.
[301] ibid 47–48, 63.
[302] ibid 46; and Finck (n 261) 25, 27–28.

public in nature, and characterized by polyarchic authority relations constituted by procedural rules'.[303] The difference between public consultations and multistakeholderism is that the latter not only engages different actors but also seeks to involve them in the decision-making.[304] In principle, multistakeholderism is posited to create a sense of ownership over regulation.[305] It could facilitate democratic deliberation, innovation, knowledge-sharing, and increased transparency.[306]

The ideal of multistakeholderism presupposes a setting where all the views are heard, debated and then fused to form a sort of synthesis, producing a compromise that gives consideration to all the viewpoints. However, its easiness again rests on the presupposed consensus over rights' meaning. Of course, this is an illusion that results in blindness toward stakeholder power relations. Behind the illusion not all stakeholders agree even on the fundamentals and not all stakeholders are created equal. Civil society organizations and scholars are often marginalized and lack genuine influence in multistakeholder tables.[307] In turn, other well-informed and abundantly resourced stakeholders, namely industry players, are strongly on the gaining side.[308] The industry dominance in these initiatives has not been left unrecognized, and after the disillusionment of consensus it may become more efficient to just sideline the cranky and market-contradicting civil society stakeholders. So, the criticism over the 'neoliberal capture' of transnational multistakeholder forums has been scathing.[309] Companies can discreetly diverge regulatory discussion to efforts that are not threatening to impede business, while still positing themselves as allies in the pursuit of the common goal of rights protection. In this sense, for the industry, multistakeholderism very much delivers its promise of regulatory ownership.

One could say that the industry capture is a problem only in (pure) self-regulation. In co-regulation, the public presence could guarantee equal participation and consideration for all the stakeholders' views.[310] However, in multistakeholder processes public regulators may have several interests at stake and the efficiency gains of industry solutions may strain their willingness to guarantee meaningful influence for other stakeholders.[311] Increased transparency can also be more high hopes than reality as procedural informality and confidentiality

[303] Mark Raymond and Laura DeNardis. 'Multistakeholderism: Anatomy of an Inchoate Global Institution' (2015) 7 International Theory 572, 573.
[304] Ben Wagner, *Global Free Expression: Governing the Boundaries of Internet Content* (Springer 2016) 161–162.
[305] Finck (n 261) 17; and Ira S Rubinstein, 'The Future of Self-Regulation Is Co-Regulation' in Evan Selinger, Jules Polonetsky and Omer Tene (eds), *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018) 509.
[306] Finck (n 261) 27; and Rubinstein (305) 509.
[307] Cohen (n 17) 228–229; and Robert Gorwa, 'The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content' (2019) 8(2) Internet Policy Review 1, 14.
[308] Agnès Callamard, 'The Human Rights Obligations of Non-State Actors' in Rikke Frank Jørgensen (ed), *Human Rights in the Age of Platforms* (MIT Press 2019) 212–213.
[309] ibid. For instance, the Global Network Initiative (GNI), as probably the most notable global multistakeholder initiative, has received plain rejections from NGOs because of the inability to provide meaningful protection. Laidlaw (n 15) 236, 238–239.
[310] Marsden (n 299) 47.
[311] Kaye (n 142) 117–118.

arguably nurture the cooperative aura of the setting.[312] In the end, this can make multistakeholderism more reminiscent of an 'invisible handshake' between regulators and industry than an open deliberative process.[313] In addition, the information asymmetry regarding the sociotechnical platform setting persists between public regulators and the companies as well.[314] Paradoxically, this is also a part of multistakeholderism's legitimacy promise: regulators *should* engage private stakeholders and stay on the back because they lack the finesse to understand the complex configurations of the Internet.[315] With such a spin, the lack of democratic accountability of multistakeholderism turns to *illegitimacy of law*. In EU context, it has been indicated that engaging stakeholders in open regulatory dialogue could fill up the democratic deficit that has long been plaguing the Union and its 'opaque legislative trilogues'.[316]

Consequently, it may be that some multistakeholder initiatives are multistakeholder only by appearances. Wagner aptly states that in transnational Internet governance more generally, multistakeholderism is effectively a 'legitimacy theatre', which is 'a form of political participation that focuses primarily on symbolic interaction'.[317] Multistakeholder conferences and other forums serve as a legitimizing front stage that hides the real workings of mainly private power.[318] Yet my claim is not that multistakeholderism *could not* deliver democratic participation and advance procedural fairness in rights protection. In practice, however, multistakeholderism may often fall short of delivering genuine regulatory capabilities for people.

3.2.2 Law and the Lack of Normative Refinement

The second outsourcing method is to impose regulatory provisions that impose generalized outcome-based obligations for companies instead of precise rules. The problem with rules is their lamentable inefficiency. Rules are rigid and thus always over- and under-inclusive.[319] It means that from time to time, rules must be changed in a slow and overly cumbersome parliamentary process. Instead, open-ended provisions allow *flexibility* for finding out the specific means for protective goals. Regulation 'can focus on outcomes rather than process, meaning that public authorities define the objectives (…) rather than precise legal rules, leaving platforms to decide how to best achieve them, encouraging flexibility and adaptability, and, providing room for maneuver to platforms'.[320] Extra leeway fosters innovation by 'harnessing a firm's own ingenuity in devising (…) solutions that meet or exceed legal requirements yet fit a firm's business model'.[321] Of course,

[312] The argument derives from the oft-invoked contradiction between efficiency and openness of the negotiations within the EU. Päivi Leino, 'Secrecy, Efficiency, Transparency in EU Negotiations: Conflicting Paradigms?' (2017) 5 Politics and Governance 6, 7–9.
[313] Michael D Birnhack and Niva Elkin-Koren, 'The Invisible Handshake: The Reemergence of the State in the Digital Environment' (2003) 8 Virginia Journal of Law and Technology 1, 18, paras 41–44. See also, Zuboff (n 190) 119.
[314] Finck (n 261) 22.
[315] Wagner (n 304) 162.
[316] Finck (n 261) 7–8, 27.
[317] Wagner (n 304) 159.
[318] ibid 159.
[319] Karen Yeung, 'Better Regulation, Administrative Sanctions and Constitutional Values' (2013) 33 Legal Studies 312, 326–327.
[320] Finck (n 261) 21.
[321] Rubinstein (n 305) 506.

on the EU level this strategy concerns mainly regulations as directives are rather principled exactly because they are to be translated to more precise rules in Member States. However, these implementation laws may only repeat the vague directive articles, as was the case with the e-Commerce Directive.

An example here is a provision that sets the so-called systemic 'duty of care', which has recently become popular in national platform regulation initiatives and on the EU level too.[322] It requires that companies scrupulously and proactively assess all the fundamental right risks of their information businesses and take adequate preventative measures to mitigate these risks.[323] Such a *carte blanche* provision fosters companies' regulatory innovation.[324] Another example is the idea of 'technology neutrality', meaning that rules regulating technologies should be so vague that they would not inhibit the deployment of still-to-come technologies. As Busch advances, 'legal rules that aim at regulating digital platforms need to be flexible enough to adapt to technological changes'.[325] Again, the flexibility avoids the inefficient and strenuous legislative process.[326]

To be sure, a systemic duty of care shifts the regulatory attention from harm liability to harm prevention and, thus, transcends the bogged-down debates on *ex post* moderation and liability.[327] In principle, such a general duty could also address the issues of amplification, manipulative micro-targeting or echo chambers. Also, open-ended norms can account for fast technological change where rules may become outdated before they've even made their way to the Official Journal.[328] Yet the daunting downside, like in rights co-regulation, is related to the fact that authoritative human rights texts contain mainly principled and open-ended norms themselves. As Fontanelli states, the application of fundamental rights requires 'normative refinement' which 'can take place at the legislative level and/or through legal interpretation and application'.[329] But rights' ideal standards are not fixed and unidirectional but constantly evolving aspirations that accommodate numerous goals in different contexts.

Even though rules are of course always imperfect generalizations, enacting legislative framework is still a place for a democratic legislator to interpret these high-level norms in a specific context. Naarttijärvi sees the requirement that law must reach certain qualitative requirements of precision and foreseeability as a facilitator of democratic deliberation: 'Maintaining qualitative requirements of legality will uphold a vital link between

[322] Natali Helberger, 'The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power' (2020) Digital Journalism 3 <https://bit.ly/3hY3akg> accessed 27 July 2020.
[323] Teresa Quintel and Carsten Ullrich, 'Self-Regulation of Fundamental Rights?: The EU Code of Conduct on Hate Speech, Related Initiatives and Beyond' in Tuomas Ojanen and Bilyana Petkova (eds), *Fundamental Rights Protection Online: The Future Regulation Of Intermediaries* (forthcoming, Edward Elgar Publishing 2020) 198–199 (in manuscript).
[324] Daphne Keller, 'Broad Consequences of a Systemic Duty of Care for Platforms' *The Center for Internet and Society and Stanford Law School* (1 June 2020) <https://stanford.io/2XkFFtG> accessed 29 June 2020.
[325] Busch (n 279) 128.
[326] Markus Naarttijärvi, 'Legality and Democratic Deliberation in Black Box Policing' (2019) Technology and Regulation 35, 45.
[327] Quintel and Ullrich (n 323) 198.
[328] van Gestel and van Lochem (n 289) 30.
[329] Filippo Fontanelli, 'The Court of Justice of the European Union and the Illusion of Balancing in Internet-Related Disputes' in Oreste Pollicino and Graziella Romeo (eds), *The Internet and Constitutional Law: The Protection of Fundamental Rights and Constitutional Adjudication in Europe* (Routledge 2016) 94.

the language of the law (which holds democratic legitimacy through the deliberations and decisions [of the parliament] that precede it) and the actions and decisions of the state'. He calls it 'qualitative legality'.[330] Importantly, qualitative legality affords also a possibility for the legislator to inject political meaning into freedom of expression and other fundamental rights. Correspondingly, if the relevant regulation only repeats the authoritative high-level standard, the power to impose the meaning of rights shifts from the legislator to those the standard-like provisions are addressed: primarily to platform companies.

Similarly, technology neutrality 'becomes in effect a transferal of power from the democratic arena to the architects of the digital arena; in some cases, this shifts power from the state to markets, in others from parliament to government agencies. In many cases it is both'.[331] As flexible norms rarely require amendments to accommodate societal changes, the interpretative development of concepts and the very meaning of rights may slip from the control of democratic legislators. Technology is embedded is social contexts and new technologies may have been designed so that they afford harming people's fundamental rights. These should be debatable in democratic forums. For instance, machine-learning algorithms are more opaque and arguably afford more threats to rights than simpler hash-matching re-upload filters used in hard moderation. Similarly, a duty of care shifts all the attention and hopes to the monitoring and enforcement of such a provision.

3.3 Administrative Framework: Cooperation in Monitoring and Enforcement

Traditionally, public administration and courts have been tasked with the contextualization of the fundamental rights obligations to real-life situations. In this regard, the full effects of outsourcing become visible in monitoring and enforcement of fundamental rights safeguards. This sub-chapter analyses how cooperation plays out in the implementation phase, i.e. monitoring and enforcement, which effectively become entangled.

3.3.1 Monitoring Regulatory Chains

According to Cappello, 'the use of co-regulatory models for platform providers leads to a shift in the interpretation of concepts that were previously assessed by the courts and authorities, such as that of harmful content, and to a "privatisation of the law" when it comes to freedom of speech'.[332] The effect is the same with overly broad binding norms: positive obligation doctrine is transferred within the platform companies, handled as an internal matter. When a duty of care and other fundamental rights responsibilities travel within the

[330] Naarttijärvi (n 326) 43. In the jurisprudence of the ECtHR, this falls within the requirement 'in accordance by the law'. *Editorial Board of Pravoye Delo and Shtekel v Ukraine* ECHR 2011–II 393, para 51: 'The law itself must correspond to certain requirements of "quality". In particular, a norm cannot be regarded as a "law" unless it is formulated with sufficient precision to enable the citizen to regulate his conduct'. See also, *Uzun v Germany* ECHR 2010–VI 1, para 61; and *Magyar Jeti Zrt v Hungary* App no 11257/16 (ECtHR, 4 December 2018), paras 58–59.
[331] Naarttijärvi (n 326) 45.
[332] Cappello (n 295) 33.

companies, they are recast managerial *compliance*.[333] Thus, the administrative framework may be to large extent substituted with managerial content moderation procedures. Waldman warns:

> [S]ome legal regimes characterized by vague requirements and process-oriented safe harbors give compliance professionals on the ground the opportunity to frame the law in accordance with managerial values like operational efficiency and reducing corporate risk rather than the substantive goals the law is meant to achieve (…) This opens the door for companies to create structures, policies, and protocols that comply with the law in name only.[334]

It is 'the paradigm of best practices'.[335] For instance, while a systemic duty of care requires internal risk assessments, risks can be assessed selectively, superficially and behind the veil of corporate secrecy. That poses immense strain for democratic accountability and substantive fulfillment of rights. More generally, translating complex fundamental rights threats to 'risks' contains many pitfalls because, as Hildebrandt notes, compressing threat to risk denotes quantification which in turn presupposes a consensus on what is quantified.[336]

Since open-ended norms allow for innovation in rights protection, the result may be something that Cohen calls 'double outsourcing'.[337] Regulatory intermediaries may devolve tasks and power to other entities, in which case those entities also become intermediaries producing 'regulatory chains'.[338] For instance, as I mentioned in Chapter 2.2, the platform companies may use independent contractors to outsource the human labor of reviewing platform content to firms like Accenture. Another instance of double outsourcing may be at hand when companies decide to protect rights through hard content moderation algorithms that detect and filter, for instance, illegal hate speech which does not enjoy the protection of freedom of expression. Algorithms disregard context completely.[339] They don't inject any meaning to fundamental rights concepts and doctrines. Instead, developers and trainers make the judicial interpretations in the (re)training of algorithms. According to Gillespie, 'it is not clear that the [judicial] human labor necessary to support such automated tools can ever go away; it can only be shifted around'.[340] Indeed, often it is shifted further down the chain and away from democratic scrutiny. While Facebook and Google largely develop filtering algorithms independently, there is a growing amount of private companies whose core business is to develop suitable automated filtering and detection tools, effectively creating a separate market for these technologies. Some

[333] Ari Ezra Waldman, 'Power, Process, and Automated Decision-making' (2019) 88 Fordham Law Review 613, 628.
[334] ibid 628.
[335] ibid 628.
[336] Hildebrandt (n 59) 77–78.
[337] Cohen (n 17) 158.
[338] Abbott, Levi-faur and Snidal (n 277) 24–25. The model of a chain is R → I → I^2 → (…) I^n → T. For simple model, see n 277.
[339] Waldman (n 333) 624.
[340] Gillespie (n 31) 107–108.

companies offer both the software and human reviewers as a necessary 'human in the loop' element for hard cases.[341]

As, in theory, these tools should be developed so that legal content protected by freedom of expression is not removed, and because I want to underscore their relevance for fundamental rights law, I call them 'balancing technologies'. Sometimes it may be more cost-efficient for a platform company to source these tools from independent developers. The third link to the chain is added when these developers outsource the training of algorithms to individual workers contracted through crowdsourcing platforms such as Amazon's Mechanical Turk.[342] Another challenge here is that privately developed tools for public interest are still proprietary, which conjures up specific challenges for public scrutiny and accountability. For instance, the workings of these technologies might be sheltered by trade secrecy.[343]

However, platform regulation also increasingly designates public administrative entities for monitoring and enforcing company obligations. In principle, this is set to guarantee at least some public control and ensure that company compliance can interrogate business models too. This is often the idea behind co-regulation or a systemic duty of care as well.[344] The catch of co-regulation is thought to be the possibility to fruitfully combine the benefits of informal flexibility but at the same time orient the conduct of the regulated toward public interest and rights.[345] But any system that tries to use private regulatory intermediaries for balancing interests is inherently unstable.[346] The managerial trap here is that successful co-regulation, although perhaps more cost-efficient than administrative law, is not necessarily cheap. Therefore, there is a constantly looming threat that co-regulation *de facto* collapses back to self-regulation and, over time, leads to 'significant devolution of regulatory authority to the private sector'.[347] The managerial trap is that more oversight increases transaction costs. The realities of social media; scale, transnationalism and complexity, only enhance the burden. Therefore, for an efficiency-oriented public regulator, there is a constant inducement to maintain low level of coordination to keep the regulatory costs down and cost-benefit calculus positive.

In my view, plausible fulfillment of positive obligations would require that regulators interrogate work in moderation centers, hash databases, algorithm training data, platform interface code, and soft moderation algorithms. Yet monitoring long regulatory chains, where intermediaries may be situated in several countries, is cumbersome and greatly raises costs. The most cost-efficient coordination of outsourcing monitors only outcomes. This is possible by quantified performance: the transformation of rights protection to aggregated

[341] ibid 108.
[342] ibid 106–107. On Mechanical Turk platform, people can register to carry out simple tasks for a tiny compensation. It's especially handy for businesses that want to outsource the training of algorithms which requires a huge workload of mechanic labeling of texts or pictures, for instance, as incitement for hatred. See 'Overview' *Amazon Mechanical Turk Inc* <www.mturk.com/> accessed 2 August 2020.
[343] Naarttijärvi (n 326) 38.
[344] Quintel and Ullrich (n 323) 199.
[345] Cappello (n 295) 3–4; Marsden (n 299) 46–47.
[346] Abbott and others, 'Competence versus Control: The Governor's Dilemma' (2019) Regulation & Governance 1, 11–14 <https://bit.ly/2BUxVat> accessed 20 May 2020.
[347] Cohen (n 17) 187.

figures and rates. Quantification transforms rights protection to simple numerical representation for speedy and efficient evaluation.[348] In the context of platforms, performance evaluation is based on transparency reports under specific transparency requirements. But scholars has identified several deficiencies in the use of simplistic indicators: reports becoming an aim in itself; the dependency on technical expertise; and 'the distortion of public values into numbers' where phenomena that are not easily measured are left unreported.[349] The figures of transparency reports are often obscure and tell little how concepts and categories are interpreted.[350] In addition, as regards business sustainability reporting in general, '[e]ven among those actors who read the reports, many do not find them useful'.[351]

In the end, the public authorities responsible for coordination might not have any idea how the imposed standards are practically fulfilled. In the end, as is often the case with outsourcing in global business, the threat is that the original outsourcing entity (EU/Member States), let alone wider public, has no effective control over the other end of the regulatory chain, where the adjudicative work is ultimately done.[352]

3.3.2 Bargaining Enforcement and Impaired Horizontality

In terms of enforcement, one could expect that the receding of legislator, informed by the change from rigid rules to flexible rarely amended norms, could usher in a golden age of principled constitutional adjudication to check that companies do not implicate users' rights while supposedly protecting them. Indeed, as rights can collide, courts normally recourse to principled and precedent-based proportionality test that balances different rights and interests by 'optimizing' the rights enjoyment case by case.[353] The quest for this contextual fair balance is at the core of constitutional doctrine and also brings on the evolvement of fundamental rights.[354] The slow process of interpretation ultimately constructs the constitutional doctrine. However, the full-blown individual court review appears unsuitable when courts are faced with the challenges of scale and complexity.[355] The Internet produces a bulk of 'low-interest' cases that are hard to manage by full-blown court review.[356] It's often of little use to challenge a single hard moderation case in a court process that may take years and costs a lot. In turn, as I mentioned in Chapter 2.3, soft moderation harms are often intangible and invisible for individuals. Consequently, few online cases reach the courtroom and even fewer reach the highest

[348] Sarfaty (n 286) 104.
[349] ibid 121–123. Similarly, Schneider (n 136) 135–137.
[350] Robert Gorwa and Timothy Garton Ash, 'Democratic Transparency in the Platform Society' in Nate Persily and Josh Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (forthcoming, CUP 2020) 17–18 <https://osf.io/preprints/socarxiv/ehcy2/> accessed 28 May 2020.
[351] Sarfaty (n 286) 118.
[352] On the troubles and costs of controlling long business supply chains, see Jodi Short, Michael W Toffel and Andrea R Hugil, 'Monitoring Global Supply Chains' (2016) 37 Strategic Management Journal 1878, 1878–1881.
[353] Alexy (n 82) 47–56.
[354] Fontanelli (n 329) 101–102.
[355] Cohen (n 17) 144. She calls them the problem of numerosity and problem of harm.
[356] Riikka Koulu, *Law, Technology and Dispute Resolution: Privatisation of Coercion* (Routledge 2019) 90, 93.

(supra)national courts, the access to which is behind a long list of admissibility criteria. Also, the fast development of technology makes reliance on precedent often hard.[357]

The challenges of court review inform two changes in the enforcement of fundamental rights safeguards. Firstly, on the individual level there is a trend to rely on speedy and efficient online dispute resolution mechanisms for redress. The provision of these mechanisms can again be outsourced to the companies. All the biggest platform already provide an appeal mechanism for hard moderation cases.[358] The procedural obscurity of online dispute resolution, however, poses challenges to public accountability and procedural fairness.[359] In addition, it is impossible for people to assert their rights claims against the companies through new low-threshold mechanisms. Companies are naturally adjudicating fair balance only between users, not between rights of the company and rights of the user(s). So, for instance, if one feels that her freedom of expression and information is implicated by a platform's potentially manipulating hypernudging, one should not expect much success from the standardized online appeal process, operated by the suspect company itself.

Secondly, specialized administrative agencies are increasingly tasked with not only monitoring but also enforcement of new standard-like norms.[360] Importantly, the expert-armed administrative agencies could invent more workable ways to interrogate the company power than courts.[361] Arguably, this includes the ability to address systemic rights issues better than court adjudication which is often more tied to the context of a specific case. In case *ESMA* (standing for European Securities and Markets Authority), the CJEU was willing to accept the devolvement of overseeing powers to the agency exactly because issues of finance are information-thick as well, and the agency can be more easily vested with the technical expertise needed.[362]

The agencies that pop up are typically vested with a wide repertoire of administrative powers. Sanctions create a continuum from rather lenient warnings and 'compliance negotiations' to highly stringent coercive measures.[363] According to Yeung, the more lenient end of the continuum exhibits 'bargaining'. It means procedurally informal negotiation between the regulator and regulated, as opposed to formal and public adjudication.[364] Often the informal process is inscrutable to the public eye and only the outcomes of enforcement processes are publicized.[365] Importantly, the more coercive sanctions are prone to years-long company challenge in court, which makes them less cost-effective for the regulator compared to the consensus-driven compliance enforcement measures.[366] The preference for more benign sanctions is in line with the

[357] Fontanelli (n 329) 99–107.
[358] OnlineCensorship.org, 'How to Appeal' <https://tinyurl.com/y48x9ezr> accessed 22 July 2020.
[359] Koulu (n 356) 87–95.
[360] Yeung (n 319) 330–338.
[361] Cohen (n 17) 200; and Abbott and others (n 346) 7.
[362] Case C-270/12 *ESMA*, EU:C:2014:18, paras 83, 85–86, 105.
[363] Both Cohen (n 17) 188–189; and Yeung (n 319) 324–325, draw the connection to the classic theory of 'responsive regulation' that constructed a pyramid of enforcement 'tools' based on their expected costs. Unsurprisingly, the most lenient and cooperative tools are also the most cost-efficient. Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate* (OUP 1992) 35–41.
[364] Yeung (n 319) 326–331.
[365] ibid 330–331.
[366] ibid 330. By contrast, the biggest platform companies have vaults of resources.

findings of Short, who has observed that from the 90s to mid-2000s, 'coercive-state anxiety' rose in legal scholarship that located the main problem of regulation in its coercive nature.[367] But as we saw in Chapter 2.1, positive obligations often limit the rights of others and thus *presume* coercion. Disproportionate attention on state coercion also produces a disparate lean toward negative obligations and the benignity of enforcement may left the fulfillment of the positives ones half-way.

Together, informality and obscurity of new monitoring and enforcement mechanisms may render even adequate oversight almost as opaque as company measures and sever them from democratic control.[368] At the same time, in the fear of populist authoritarian rise, EU policies have sometimes tended to stress rule of law as requiring independence from the government.[369] However, the threat to independence may come from two directions. The opaqueness of monitoring/enforcement and the presupposed consensus over fundamental rights can make the regulatory agencies prone to capture. In regulatory capture, 'intermediaries become the leaders in regulation, with the ostensible rule-makers following them'.[370]

In sum, regulatory cooperation gives rise to something I call 'impaired horizontality'. On the one hand, regulation does indeed recognize that the relationships of social media users, as private actors, must be regulated by installing a legislative and administrative framework on social media. On the other hand, however, to control people's behavior more effectively and efficiently, the cooperative framework devolves positive obligations to regulatory intermediaries, i.e. *within* the platform companies, from which they may be further diffused to regulatory chains that are highly resistant to democratic scrutiny. As we saw in Chapter 2.1, positive obligations are indeed hard to define acontextually. That's why the ECtHR often leaves the determination of the most suitable means for their fulfillment to the discretion of the state under the doctrine of margin of appreciation.[371] Outsourcing through standards, coupled with administrative oversight and informal bargaining makes it also harder to discern public interference from private and other way around. In sum, state-mediated horizontality becomes only lightly mediated by EU/Member States.

Moreover, cooperative framework threatens to enhance company power instead of scrutinizing it. The companies have perhaps become more amenable for positive obligations that require maintaining safe public sphere, for instance, by sanctioning harmful expressions or stamping fact-checking labels.[372] However, they cannot critically assess the most striking power asymmetry on social media that may threaten users' free expression, namely the power relationship between themselves and users. No 'best practice' devised by the companies can undermine the companies' own information organization business, even if the business would be the source of abuse. Therefore, the companies are incapable of implementing the non-domination principle

[367] Jodi L Short, 'Paranoid Style in Regulatory Reform' (2011) 63 Hastings Law Journal 633, 668–673.
[368] Cohen (n 17) 234.
[369] European Parliamentary Research Service (n 268) 10–11.
[370] Abbott, Levi-faur and Snidal (n 277) 30.
[371] Eg, *Söderman v Sweden* ECHR 2013–VI 203, para 79. It is the same with the CJEU, Beijer (n 79) 306.
[372] Rikke Frank Jørgensen, 'Framing Human Rights: Exploring Storytelling within Internet Companies' (2018) 21 Information, Communication & Society 340, 349–350.

to the fundamental rights framework. This job would fall to the public administration. Unfortunately, the very idea of cooperation between public and private power prefers consensus and discourages outright challenge. Lastly, even though the CJEU has recently embraced horizontality of the Charter more candidly,[373] for the reasons indicated above, traditional *ex post* individual court review does not sit well with content moderation issues and complex technology. This makes it harder for courts to check the legally enhanced power of the companies, rendering the doctrines of indirect (and direct) horizontality less effective too. Thus, impaired horizontality has great challenges in fostering people's democratic self-determination.

3.4 EU Contest: Fiat Interdiction

Despite the seemingly harmonious cooperation in the protection of users' freedom of expression and other rights, the EU and the platform companies have also crossing aims.[374] The previous sub-chapters outlined the cooperative part of regulation which is facilitated by the presupposed consensus over the meaning of free expression (and related rights). But regulation could not come into being if the EU would not observe that company conduct leaves to be desired. However, we must remember that neither the EU/Member States nor the companies are usually positioned to unilaterally dictate the regulation, which results in necessary compromises that fuse contest with cooperation. This sub-chapter analyses the element of contest by the EU. In this sub-chapter it becomes clear that the EU has chosen to advance also less benign techniques. This has led the companies to pick up free expression arguments to challenge these claims, which is explored in the next sub-chapter.

The EU legislator has many vital interests at stake in content regulation. According to Cohen, in some cases, the regulator is seeking to assert its sovereignty over platform companies to dictate information flows unilaterally.[375] Government measures deploy the logics of 'fiat interdiction', which denotes draconian impositions for companies to block forbidden information. The mandates designate specific types of content as existential threats for public interests. Thus, urgent and exceptional countermeasures are purportedly justified.[376] One could assume, for instance, that the proprietary algorithms affording wide-scale manipulation for political-economic gain warrant exceptional regulatory measures. But the monsters sketched here are more conventional and potentially more suspect too. Within tech circles in the US, the threats were called 'the "Four Horsemen of the Infocalypse": terrorism, drug dealers, pedophiles, and organized crime' with large-scale IPR infringement appearing soon after.[377] In Europe, especially terrorism and threats to IPR have emerged as existential threats that require exceptional public interference.[378]

[373] See generally, Eleni Frantziou, '(Most of) the Charter of Fundamental Rights is Horizontally Applicable' (2019) 15 European Constitutional Law Review 306.
[374] Abbott, Levi-faur and Snidal (n 277) 30, noting that intermediaries' goals are not always in the public interest.
[375] Cohen (n 17) 109–110.
[376] ibid 110–111.
[377] ibid 110.
[378] As regards IPR, it is illustrating that before 2019 the CJEU jurisprudence on platform liability under the e-Commerce Directive revolved solely around IPR violations. See eg, Case C-324/09 *L'Oréal v eBay*; and Case C-360/10 *SABAM v Netlog*.

Therefore, often the logic of fiat interdiction brings regulation into existence in the first place. However, the lack of effective control, coupled with the need for utilitarian efficiency, may reduce fiat interdiction to cooperation. For instance, in Chapter 3.3.2, I argued that administrative agencies oversee platforms with preference for cost-efficient means of benign compliance bargaining. Yet we noticed that broad administrative mandates afford more coercive powers as well. Sometimes public incursions surmount the intensive anxiety over costs, producing regulation that forfeits cooperation and seeks to control users with raw power that formally resembles traditional 'command-and-control' regulation.[379] Usually this is manifested in straightforward orders to take down individual pieces of content that a public authority has either generally or individually stamped illegal. Again, characteristic here are justifications for draconian powers of law enforcement that invoke the language of extraordinary threats.[380] However, I have already argued that public oversight is not always much more democratically accountable than private governance. Especially the operations of police and intelligence agencies may be veiled in ominous secrecy that seeks to ensure effectiveness/efficiency.[381] So, for instance, it was recently claimed that the French Internet Referral Unit had demanded a removal of 550 URLs from the Internet Archive which contained publications from governments among others. These were arguably falsely stamped as terrorist propaganda.[382]

Thus, the logic of fiat interdiction of course raises the familiar worries of government overreach and even censorship. Of course, in many countries there is a real reason to worry for government overreach, as many authoritarian regimes have been eager to claim their sovereignty over platforms to suppress political opposition, minorities, or human rights activists.[383] For instance, Gillespie recounts how Russia has been particularly innovative in curtailing 'forbidden information', grounding its justifications in a wide array of existential threats ranging from cybercrime to surveillance by the US National Security Agency.[384] However, censorship fears have also been raised against EU's recent interdiction efforts.[385] Yet at least in well-functioning democracies, fiat interdiction is harder to impose without constitutional restraint simply because the government overreach is something that constitutional law is accustomed to handle with the straightforward doctrine of negative obligations. So, for instance, it is unquestionable that law enforcement is, at least in principle, always obliged by fundamental rights and courts are familiar with scrutinizing state practices. For instance, the ECtHR is steadily mounting a pile of case-law on different government incursions, interrogating administrative powers without proper fundamental rights safeguards.[386]

[379] For the concept of command-and-control regulation, see Short (n 367) 658–662.
[380] Cohen (n 17) 120.
[381] Naarttijärvi (n 326) 38.
[382] Chris Butler, 'Official EU agencies falsely report more than 550 Archive.org URLs as terrorist content' *Internet Archive* (10 April 2020) <https://bit.ly/2EFE0Zf> accessed 26 June 2020.
[383] Kaye (n 142) 101–111.
[384] Gillespie (n 31) 38–39.
[385] Citron (n 149) 1070.
[386] See eg, *Ivashchenko v Russia* App no 61064/10 (ECtHR, 13 February 2018), paras 81, 89–93; and *Big Brother Watch and Others v the United Kingdom* App nos 58170/13, 62322/14 and 24960/15 (ECtHR, 13 September 2018), paras 487–495; and *Roman Zakharov v Russia* ECHR 2015–VIII 205, paras 298–300, 302–304.

3.5 Company Contest: Free Expression Management

Regulatory intermediaries normally have their own diverging interests they seek to advance with their new legally endowed powers.[387] Indeed, the platform companies have developed their own strategies to contest EU's unilaterally imposed mandates of fiat interdiction that may not afford the same leeway for the companies as regulatory cooperation. Cohen gathers them under the logics of 'innovative and expressive immunity' which includes strategies both in the courtroom and within the realm of public relations. I name one of these immunities 'free expression management'. This sub-chapter outlines the elements of free expression management that companies mount to challenge the EU and European regulators.

3.5.1 Platform Companies as Fundamental Right Subjects

There is extensive scholarship on how the platform companies have triggered the First Amendment protection to shelter their business operations from regulatory scrutiny.[388] However, while in Europe companies are also entitled to free expression,[389] it is doubtful that social media companies' hard moderation systems or algorithmic soft moderation can be considered companies' exercise of free expression. The picture is slightly more complex. Under the ECtHR jurisprudence, platforms as media enjoy indirect free expression protection as facilitators of *peoples'* expressions.[390] Consequently, the platform companies seek to place themselves at the forefront in struggles over users' free expression and invoke 'narratives about heroic civil libertarian opposition to state censorship'.[391] While the idea of privately protected freedom of expression challenges the logic of fiat interdiction outlined above, it is entirely in line with neoliberal governmentality. As with regulatory cooperation, free expression management advances again the proposition that positive obligations flowing from freedom of expression are well fulfilled by private companies.

Freedom of expression and information has for long been fused in the companies' political economy. For instance, freedom of expression is inextricable part of the companies' outspoken missions. YouTube's four values are freedom of expression, freedom of information, freedom of opportunity, and freedom to belong.[392] Twitter in turn is promoting free expression and 'protecting the health of the public conversation around the world'.[393] Jørgensen has conducted insightful qualitative research on how freedom of expression is framed within Facebook and Google. It reveals that the companies have indeed deeply internalized their new role as frontline fighters.[394] Yet, as Schwarz points out, private governance is not even nominally impartial.[395] In

[387] Abbott, Levi-faur and Snidal (n 277) 26, 30.
[388] See eg, Zuboff (n 190) 108–112; Keller (n 273) 17–20; Cohen (n 17) 93–97.
[389] Khoury and White (n 87) 142.
[390] See eg, *Jersild v Denmark* App no 15890/89 (ECtHR, 23 September 1994), para 31: 'Article 10 protects not only the substance of the ideas and information expressed, but also the form in which they are conveyed'. See also, Dobber, Fathaigh and Borgesius (n 216) 8–10.
[391] Cohen (n 17) 135.
[392] YouTube, 'About' *YouTube LLC* (2020) <www.youtube.com/intl/en-GB/about/> accessed 7 March 2020.
[393] Twitter, 'Our Advocacy' *Twitter Inc* (2020) <https://tinyurl.com/y256pwwa> accessed 7 March 2020.
[394] Rikke Frank Jørgensen, 'Rights Talk: In the Kingdom of Online Giants' in Rikke Frank Jørgensen (ed), *Human Rights in the Age of Platforms* (MIT Press 2019) 167–168, 171; and Jørgensen (n 372) 344–350.
[395] Schwarz (n 34) 122.

effect, free expression management seeks to impose a very particular framing for freedom of expression that works in the political-economic interests of companies.[396] This framing mandates vigilance when it comes to threats from states but is willfully blind to any issue that would stem directly from the companies' businesses. Cohen notes that although platform companies' resistance against public incursion may 'also implicate users' rights of privacy, expression, and association, platforms more often seem to be principally concerned with establishing their own regulatory independence'.[397] Suzor claims that '[u]ser interests have been able to be protected when they align with the interests of the tech sector. But often, technology companies are most worried about their own legal risks and costs'.[398]

If we take free expression management as a strategy to maintain information sovereignty, it becomes visible that it's not confined to freedom of expression. Companies' self-serving frame of free expression receives backing from the fact that companies' businesses of information sovereignty are protected by freedom to conduct business, as enshrined in Article 16 of the Charter, and the right to property, as enshrined in Article 17.[399] Kaye insightfully states: 'It remains an open question how freedom of expression concerns raised by design and engineering choices should be reconciled with the freedom of private entities to design and customize their platforms as they choose'.[400] For instance, in the European case law, the Internet companies' right to conduct business has sometimes played a more significant role than freedom of expression when the CJEU has struck down the mandates to monitor unlawful content.[401]

3.5.2 Free Expression Protection vs Free Expression Management

In my view, free expression management is rooted in three intertwined assumptions on the right's meaning that frame the scope of protection, i.e. what harms are recognized under freedom of expression and what are left out. These three assumptions differentiate free expression protection, which aims to nurture democratic self-determination, from company-served free expression management. These assumptions are: 1) that freedom of expression and information is connected to presumptive virtuousness of unimpeded private innovation; 2) that freedom of expression is fulfilled through individual user 'empowerment'; and 3) that freedom of expression is meant to protect primarily against public incursions. In turn, freedom of expression

[396] Abbott and others remark that intermediaries can informally interpret rules to make them more amenable for their interests. Abbott, Levi-faur and Snidal (n 277) 30.
[397] Cohen (n 17) 236.
[398] Suzor (n 169) 156.
[399] Right to property is enshrined also under the ECHR in Article 1 of Protocol 1 (as ameded by Protocol 11). While there is no obvious parallel to Article 16 of the Charter in the ECHR, the ECtHR has afforded protection for business operations through the right to property and sometimes the right to freedom of expression. European Union Agency for Fundamental Rights, *Freedom to Conduct a Business: Exploring the Dimensions of a Fundamental Right* (Publications Office of the European Union 2015) 10.
[400] UNCHR, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (11 May 2016) UN Doc A/HRC/32/38, 15.
[401] Case C-360/10 *SABAM v Netlog*, EU:C:2012:85, para 47. Similarly, in *Facebook Ireland* CJEU ruled that monitoring obligations were lawful only when the companies implement them through automated technologies that are more cost-efficient in long term. Case C-18/18 *Facebook Ireland Limited*, EU:C:2019:821, para 46.

protection could potentially offer protection beyond these imperatives to account for harms that don't fall into the frame of management.

Firstly, in informational capitalism innovation rises to prominence as it becomes the crucial competitive weapon in the market.[402] Its presumed goodness is connected to the Silicon Valley efficiency ideology prescribing that social and political issues are solvable through Internet technology. Therefore, the ongoing parade of new applications and other tools would eventually lead us to a some sort of technosolutionist utopia.[403] Indeed, 'freedom of expression is described as inherent in the services themselves' turning the right into 'product features'.[404] This ideology informs, for example, Facebook's political-economic statements that frame its services as ubiquitous, foundational, if not essential for being social.[405] Moreover, the innovation imperative sees regulation as its 'mortal enemy', from which it follows that regulation that constrains free experimentation is to be resisted with full force.[406]

In reality, innovation is not inherently good, but can sometimes be harmful and dangerous.[407] All the mechanisms of platforms have certainly been innovative. While they've brought many benefits that are well-known, the same innovations have also brought new harms and amplified many existing ones, as we explored in Chapter 2. While some of these consequences have likely been unintended, others have not. For instance, it has been recently revealed that Facebook had indeed understood that its innovative soft moderation algorithms were not connecting people, but in fact driving them apart, and that some of its algorithms were disproportionately implicating racial minorities. Yet the company pushed ahead with harmful innovation.[408] These choices are understandable when we remember that private innovation is self-interested. As any innovated solution must fit to the business model, the political economy of the companies imposes necessary boundaries for their innovative capabilities. Indeed, it has been observed that technoscientific innovation has become thoroughly marged with *data rentiership* where 'people and businesses are innovating in order to create data assets and data rents: they come up with new mechanisms, devices, or instruments of data

[402] Castells (n 285) 188.
[403] Morozov (n 213) x–xiv, 5–9. Morozov sarcastically notes that Silicon Valley's prime slogan has changed from 'Innovate or Die!' to 'Ameliorate or Die!'. ibid x.
[404] Jørgensen (n 372) 346.
[405] David B Nieborg and Anne Helmond, 'The Political Economy of Facebook's Platformization in the Mobile Ecosystem: Facebook Messenger as a Platform Instance' (2019) 41 Media, Culture & Society 196, 199, 207; and Anna Lauren Hoffmann, Nicholas Proferes and Michael Zimmer, '"Making the World More Open and Connected": Mark Zuckerberg and the Discursive Construction of Facebook and Its Users' (2018) 20 New Media & Society 199, 204–206.
[406] Cohen (n 17) 178; and Zuboff (n 190) 104.
[407] Cohen (n 17) 90–93; and Morozov (n 213) 167–168. The spin here, as Morozov states, is that 'innovations that fail or lead to disastrous results are naturally not considered part of the innovation vocabulary; technologies are innovative only if they are successful and risk-free' (168).
[408] Jeff Horwitz and Deepa Seetharaman, 'Facebook executives shut down efforts to make the site less divisive' *Washington Post* (26 May 2020) <https://on.wsj.com/2PdX4js> accessed 23 July 2020; and Olivia Solon, 'Facebook ignored racial bias research, employees say' *NBC News* (23 July 2020) <https://nbcnews.to/3fg9cLm> accessed 29 July 2020.

ownership and control that are designed to extract value through the ownership and control of personal data'.[409] So, there is a striking mismatch between the front that posits companies as altruistic 'connectors of the world' in the name of free expression and information, and the practical backside quest for the appropriation of data.

The second assumption of the individual empowerment cherishes the very traditional liberal ideal of the rational individual who dispassionately assesses the quality of information and disregards demagogy.[410] For Silicon Valley, this self-reliant rational individual ideal is fused with the idea of consumer who is always right.[411] For YouTube/Google, user empowerment is afforded by their tools and it means 'choice, opportunity, and exposure to a diversity of opinion'.[412] This leads to despise for the editors of old media that don't promote consumer choice but passive media consumption.[413] Decentralized social networks afford interaction. Thus, it follows that more information means more choice, and freedom of information equates with 'the more the better'. In addition, as rational deliberation requires information as a basis for decision-making, more information purportedly results in more rationalism.[414] But as Pollicino points out, people's time has not increased from 24 hours a day.[415] Neither have their capabilities for maintaining attention reached super-human levels. Social networks do not only replace old media gatekeepers but also place social media algorithms as new power hubs of information control.[416] As platforms are characterized by infoglut, users become highly dependent on platform features and recommendations that can make the glut legible.

In reality, a platform does not even aim to set the stage for users' rational deliberation but instead to structure information so that it *nudges* people toward specific choices that, while marketed as consumerist personalization, are often motivated by political-economic interests. Similarly, the interfaces encourage habitual and mind-numbing clicking of stylized 'like' and other social buttons instead of conscious interaction.[417] Of course, nudging does not lead to guaranteed user choices – that would mean belittlement of agency. Nevertheless, it's not self-evident that social media leads to significantly better 'user empowerment' than linear media, and one cannot exactly describe it leading to individual rationality unbound. Incessant nudging reveals a paradox when it's coupled with the industry's empowerment rhetoric: neoliberal individuality is celebrated while more realistic, context-dependent individual agency is undermined by manipulative nudging. Moreover, one-sided imperative of individual personalization does not consider that

[409] Kean Birch, Margaret Chiappetta and Anna Artyushina, 'The Problem *of* Innovation in Technoscientific Capitalism: Data *Rentiership* and the Policy Implications of Turning Personal Digital Data into a Private Asset' (in press, 2020) Policy Studies 8 <https://bit.ly/3hWpOcP> accessed 23 July 2020.
[410] Ananny (n 223) 7–9; and Cohen (n 207) 6–10.
[411] Morozov (n 213) 148. More generally, consumer choice is the crown jewel of neoliberalism. Grewal and Purdy (n 53) 13.
[412] Juniper Downs, Google Senior Policy Counsel, as cited in Kaye (n 142) 42.
[413] Morozov (n 213) 145–167, 173–180.
[414] ibid 86–89.
[415] Pollicino (n 179) 30 (in manuscript).
[416] van Dijck, Poell and de Waal (n 133) 40–41.
[417] Cohen (n 207) 652, 655.

maybe the relativization of all information may not only erode the existing societal basis, but also make it harder for people to come together to build a new one.

From the two previous assumption logically follows the third that sees the role of the government mainly through negative obligations. Jørgensen summarizes: 'As such, the human rights discourse is not foreign to these companies; however, it has been firmly rooted in the context in which governments make more or less legitimate requests and companies respond and push back against such requests'.[418] Company experts are incredulous towards government take-down requests and eagerly conduct secondary reviews against freedom of expression standards.[419] Moreover, in some cases, the companies even directly challenge government take-down orders in courts and making their case relying on freedom of expression.[420] For instance, Twitter challenges court orders mandating removal of content based on its free expression commitment.[421]

Isolated fights against states are published in the companies' transparency reports. But transparency can obfuscate also intentionally.[422] So, companies have designated reports both as a new commitment of cooperation and, *at the same time,* a discreet challenge to governments.[423] In that way, the ongoing reporting requirements we addressed in Chapter 3.3 can also be weaponized and directed back against European regulators. The number and details of every government order and request are diligently documented and publicized in transparency reports.[424] As another example, Facebook fancies publicizing reports on how it shelters fragile democracies from disinformation by foreign governments.[425] However, abrupt silence ensues when one asks about their proprietary soft moderation algorithms or advert personalization tools.[426]

Of course, independent (i.e. privately controlled) media has always been a corner stone mechanism for critiquing those in power. However, Cohen sees that the assumption that regulating speech equates with censorship has created a hysteria that impairs regulators from addressing the novel and still poorly understood harms of social media.[427] The looming threat of censorship arguably informs the doctrine of positive obligations on freedom of expression too. Kuczerawy points out that positive obligations are in general less

[418] Frank Rikke Jørgensen, 'When Private Actors Govern Human Rights' in Ben Wagner, Mattias C Kettemann and Kilian Vieth, *Human Rights and Digital Technology* (Edward Elgar Publishing 2019) 352.
[419] Jørgensen, 'Rights Talk' (n 394) 167–168, 171; and Jørgensen, 'Framing Human Rights' (n 372) 349–350.
[420] Callamard (n 308) 213.
[421] Twitter, 'Defending and respecting the rights of people using our service' *Twitter Inc* (2020) <https://bit.ly/33j7syH> accessed 17 April 2020. One example states: 'We objected to a court order related to two accounts and two Tweets posted by the two accounts, for causing harm to the reputation and credibility of one of the largest banks of Turkey. Our objection was partially accepted and the initial court decision was cancelled due to freedom of expression.' Twitter Transparency Report, 'Removal Requests – January to June 2019: No Action' *Twitter Inc* (2020) <https://transparency.twitter.com/en/removal-requests.html> accessed 17 April 2020.
[422] Ananny and Crawford (n 255) 979.
[423] Cohen (n 17) 135.
[424] Jørgensen (n 394) 171. See eg, Facebook Transparency, 'Content Restrictions Based on Local Law: Jan – Jun 2019' <https://transparency.facebook.com/content-restrictions> accessed 17 April 2020.
[425] Eg, Nathaniel Gleicher, 'Removing more coordinated inauthentic behavior from Iran' *Facebook Newsroom* (28 May 2019) <https://about.fb.com/news/2019/05/removing-more-cib-from-iran/> accessed 23 July 2020.
[426] Cohen (n 17) 136.
[427] ibid 104–105.

developed regarding freedom of expression than, for instance, the right to privacy in Article 8 of the ECHR.[428] However, companies are not against positive rights obligations that call for assertive action *per se*. For instance, according to the companies, their hard content moderation nurtures safe public sphere for all expressions.[429] Instead, platforms idea of positive obligations is distinctly selective. This idea is conveyed to regulators by implying that only through cooperation with companies they can avoid the slippery slope of censorship.

This brings us back to the concept of impaired horizontality I introduced in Chapter 3.3.2: Companies are regulating the relationships between users (as private actors), for instance, when they regulate hate speech to maintain a public sphere that fosters everyone's right to express themselves. At the same time, however, this horizontality is impaired because it's incapable of scrutinizing the most profound power asymmetry between two private actors, namely the relationship between user(s) and a social media company. The inability persists because company's own interests will prevail in case there exists an interest conflict.[430] To be sure, it is not always that companies are playing down issues of their business with bad intentions, but within the companies people consider it *inconceivable* that their business or the algorithmic processes of platform itself could actually jeopardize free expression or other rights.[431]

The ideology that comprises the three assumptions of private innovation, individual empowerment, and state non-interference, is squarely illustrated by the distinctly American 'marketplace of ideas' metaphor.[432] It is the convergence point of market efficiency that should be free from public interference, and freedom of expression that preferably should be left with modest public interference too. While scholarship has questioned the metaphor and noted its unsuitability with European freedom of expression,[433] it is faithfully followed by platform companies and thus far from irrelevant. The reason for contesting fiat interdiction specifically with freedom of expression is the right's argumentative weight. Kaye summarizes this well by stating that:

> It is much less convincing to say to authoritarians, "We cannot take down that content because that would be inconsistent with our rules," than it is to say, "Taking down that content would be inconsistent with the international human rights our users enjoy and to which your government is obligated to uphold".[434]

The strategy works equally well against regulatory efforts that would target the algorithmic mediation of information, after all, free expression in principle protects the medium as well.[435] While Kaye is willing to rely on the companies mainly against authoritarian governments, some scholars have already been willing to throw their faith in platform companies against incursive European regulation. For instance, Citron remarks:

[428] Kuczerawy (n 22) 162. Similarly, Woods (n 1) 112–114, sees that the threshold for positive obligations of free expression is high.
[429] Jørgensen (n 372) 349–351.
[430] Lynskey (n 48) 17.
[431] Jørgensen, 'Rights Talk' (n 394) 173; and Jørgensen, 'Framing Human Rights' (n 372) 352.
[432] Pollicino (n 179) 29–32 (in manuscript).
[433] Eg, ibid 31 (in manuscript).
[434] Kaye (n 142) 119.
[435] See n 390.

'Ultimately, Silicon Valley may be our best protection against [EU] censorship creep'.[436] Others have envisioned models for companies to find rights' fair balance algorithmically.[437] To be sure, the company efforts against censoring can sometimes be laudable. However, well-meaning encouraging of company contest may look aside their interests and the undemocratic underpinnings of private protection.

4. Case Studies of Cooperation and Contest

This Chapter builds on the previous chapter to contextualize cooperation and contest by analyzing the level of hybridity in regulatory instruments. I have opted for three initiatives the EU has introduced during the last five years to analyze them as case studies. While there is also other regulation regarding social media, I have chosen these firstly because they are different in nature: one is a formally non-binding code of conduct, the second is a directive, and the last a regulation which is directly applicable. This allows me to assess how the more theoretical arguments on cooperation and contest are visible in different regulation. I can assess how the choice of instrument may affect the degree of involvement of private actors, the hypothesis being that the level of hybridity is not the same in a directly applicable regulation as in a code of conduct. I believe this may bring more insights than, for instance, assessing three different codes of conduct would.

Secondly, the material scope of the three case studies, while to some extent overlapping, is also different in that one considers primarily the prevention of hate speech, one protection of minors, and one counter-terrorism. This widens the scope compared to three instruments that would address for instance only terrorism. However, the case studies do of course converge in that all seek to explicitly give consideration to freedom of expression by maintaining a safe public sphere while also providing free expression safeguards (together forming a legislative and administrative framework). Naturally, one could say that the relevance of freedom of expression flows already from the fact the instruments are regulating media.

4.1 Code of Conduct on Hate Speech

The first case study to be examined through cooperation and contest is EU Code of Conduct on Hate Speech (hereafter the Code).[438] As the title indicates, it aims to target hate speech on the biggest social media platforms. Agreed in 2016, it is also chronologically one of the first instruments in the 'new wave' of EU regulation on social media platforms that has emerged during the last five years. Arguably, it is also the 'softest' of the instruments examined here. One could even claim that it is not content regulation, but a plain self-regulatory effort because, after all, the Code is supposedly 'voluntary commitment'.[439] However, as we saw in Chapter

[436] Citron (n 149) 1062.
[437] Mart Susi, 'The Internet Balancing Formula' (2019) 25 European Law Journal 198, 198–199.
[438] Code of conduct on countering illegal hate speech online (hereafter in footnotes Code of conduct) (2016) <https://bit.ly/3i3Jzzh> accessed 22 May 2020.
[439] European Commission, 'Code of Conduct on countering illegal hate speech online: Questions and answers on the fourth evaluation' MEMO/19/806.

3.2.1, the Commission's use of voluntariness is primarily strategic and as we will see, the Commission's degree of involvement indeed suggests otherwise.

4.1.1 EU Internet Forum

The Code is a brainchild of the multistakeholder EU Internet Forum (hereafter the Forum). The Forum was initially envisioned in the Commission's Communication on Preventing Radicalisation to Terrorism and Violent Extremism in 2014 to '[c]ooperate more closely with civil society and the private sector to address challenges faced online'.[440] The Commission would set up a forum with key industry players to discuss a wide range of issues related to 'dangerous propaganda' online. Officially, the Forum was launched on 3 December 2015 and has worked as a facilitator of cooperation ever since.[441] Despite a somewhat generic name, it has only two aims, to prevent terrorism and hate speech on Internet platforms. However, it seems that these goals have always been intertwined. Of course, with indications of the rising extremism in Europe, this link is not farfetched. In late 2016, the Commission informed that '[t]he EU Internet Forum has two key objectives: to reduce accessibility to terrorist content online; and to empower civil society partners to increase the volume of effective alternative narratives online'.[442]

Even though the Forum was to bring together multiple types of stakeholders, already at the beginning it was clear that the priority was cooperation with the industry. By contrast, civil society was envisioned to be structurally skewed from the Forum to subsumed under the parallel Radicalisation Awareness Network (RAN) initiative.[443] The launch meeting minutes picture a slightly paternalistic approach: platform companies and public stakeholders would work together to 'empower civil society partners to challenge the extremist narrative online'.[444] Yet civil society empowerment through *participation at the meetings* was not advocated. While the intention seemed to be that civil society would be allowed to the workshops of the Forum 'on a case-by-case basis',[445] it was only after several demands that NGOs European Digital Rights (EDRi) and AccessNow were allowed to the discussion tables on hate speech.[446] The European Parliament is not part of the Forum.

---

[440] European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Preventing radicalisation to terrorism and violent extremism: strengthening the EU's response' COM (2013) 941 final, 8.
[441] European Commission Press Release, 'EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online' IP/15/6243.
[442] European Commission Press Release, 'EU Internet Forum: A major step forward in curbing terrorist content on the internet' IP/16/4328.
[443] COM (2013) 941 final, 8–9. Instead, RAN is effectively handling the second Forum objective of producing 'effective alternative narratives online'.
[444] European Commission Directorate-General Migration and Home Affairs, 'Launch of EU Internet Forum: Summary of Meeting' (Redacted) Ref Ares(2016)1158809 (7 March 2016) 1 <https://bit.ly/3gnrQlU> accessed 22 May 2020.
[445] European Commission, 'Launch of the EU Internet Forum Brussels, 3 December 2015' (Redacted), Ref Ares(2016)1160160 - 07/03/2016, 1 <https://bit.ly/314bsjy> accessed 22 May 2020.
[446] EDRi, 'EU Internet Forum against terrorist content and hate speech online: Document pool' (10 March 2016) <https://edri.org/eu-internet-forum-document-pool/> accessed 22 May 2020.

The organization of the Forum comprises multistakeholder meetings on three levels, with *ad hoc* trainings, workshops and awareness raising events at the 'intermediate level' constituting the main work.[447] The workshops are meant to 'address the practical implementation of the common objectives and related actions and advis[ing] the Forum on other relevant initiatives, including in the area of hate speech online'.[448] The annual High level event engages also Member States, while the governance level would gather together at 'expert level' for follow-up and reporting activities.[449] The Forum has by now become well-known for its notorious secrecy. Barely anything has been publicly disclosed from its meetings and the influence of different stakeholders to the final text of the Code is unknown. The Commission's unwillingness to disclose information on the companies' views at the Forum, even after an access to documents request, escalated to a point where EDRi issued a complaint to the European Ombudsman.[450]

From the Commission's initial dismissing decision and the Ombudsman's decision we can observe the reasoning underlying the Forum's secrecy. It was broadly based on two justifications which can be neatly subsumed under the need for efficiency. In accordance with the logics of fiat interdiction, the first evokes a narrative of public security being under the existential threat of terrorism. The Commission noted that '[i]dentifying companies and revealing their views and action taken against terrorist content on the internet would seriously undermine the effectiveness and success of the measures implemented' and 'would reveal vulnerabilities that terrorists could further exploit'.[451] In effect, 'the prevailing interest is to secure the effective organisation and follow-up of the EU Internet Forum and ultimately the Commission's efforts to address violent extremism'.[452] The second justification privileges companies' commercial interests over the public accountability concerns. The Secretary-General summarized:

> Nor have I, based on my own analysis, been able to identify any elements capable of demonstrating the existence of a public interest that would override the need to protect the Commission's decision-making process and the companies' commercial interests in the framework of the EU Internet Forum.

This squarely sums up the pitfalls of regulatory cooperation that foresees convergence and strong interdependency between companies and the EU. It also exhibits the efficiency/democracy trade-off. The

[447] European Commission DG Home Affairs, 'Jour Fixe Meeting: EU Internet Forum' (Redacted) (10 June 2015) 3 <https://bit.ly/2Db1zZC> accessed 22 May 2020.
[448] European Commission, 'Launch of the EU Internet Forum Brussels, 3 December 2015' (n 445) 1.
[449] DG Migration and Home Affairs, 'Jour Fixe Meeting: EU Internet Forum' (n 447) 3. Such a statement of course indicates the presupposed consensus that I described as a foundational element of regulatory cooperation.
[450] Kirsten Fiedler (EDRi), 'Complaint about maladministration' (17 February 2016) <https://bit.ly/2Dohzaq> accessed 22 May 2020.
[451] European Commission Secretary General, Decision of the Secretary General on behalf of the Commission pursuant to Article 4 of the implementing rules to Regulation (EC) N° 1049/20011, C (2016) 2065 final, 6 <https://bit.ly/3fgbJFm> accessed 22 May 2020.
[452] ibid 7–8. The Secretary-General hypothesized rather wild threats that the disclosure of companies' views would expose their representatives to terrorist attacks themselves, despite the fact that these representatives are notable public figures who have expressed their views publicly in several contexts, including the Forum press releases. See eg, Bickert and Fishman (n 156); and IP/16/4328.

Ombudsman Emily O'Reilly was convinced by the Commission's arguments with barely a hint of critique.[453] Being gripped by the existential threat of terrorism, she was effectively also protecting platform companies' right to business at the expense of public's capabilities to scrutinize regulation.

While we cannot really know who contributed to what part of the Code and how, it has become clear that the influence of the civil society stakeholders was minimal. When the Code was released, EDRi and AccessNow roundly rejected its statements, announced their thorough sidelining during the discussions, and resigned from the Forum altogether.[454] Thus, we can see that the gap between the ideal of co-regulatory multistakeholderism and the practice could hardly be any wider. The Commission seemed so anxious to see the Forum 'succeed' that it felt the trade-off necessary. Fruitful cooperation purportedly requires that the informal setting of discussions is nurtured by opaqueness. Indeed, the Commission reminded EDRi that

> engagement with different stakeholders and in particular with the industry is based on a relationship of mutual trust among all stakeholders involved. Public disclosure of the positions or proposals of stakeholders or Member States would undermine that necessary climate of mutual trust.[455]

One may wonder, where exactly lies the extra openness compared, for instance, to the ordinary legislative process so lambasted by the proponents of multistakeholderism.[456] As the above analysis shows, public facilitation does not automatically guarantee equal consideration for all the stakeholders.

4.1.2 Code and Its Monitoring

After analyzing the process preceding the Code, we can proceed to its content. The document is somewhat brief. It is divided into a more introductory part and an ensuing list of 12 more specific commitments. The introductory part starts with a solemn proclamation:

> Facebook, Microsoft, Twitter and YouTube (…) – also involved in the EU Internet Forum – share, together with other platforms and social media companies, a collective responsibility and pride in promoting and facilitating freedom of expression throughout the online world.[457]

However, straight after it is noted that the companies also share the Commission's and the Member States' commitment to fight against illegal hate speech, as defined in Article 1 of the Council Framework Decision 2008/913/JHA.[458] Notably, the Parliament is not mentioned, as it was not invited to the Forum in the first place. The purpose of the Code is primarily to serve the companies by 'guiding their own activities as well as sharing

[453] European Ombudsman, Decision in case 292/2016/AMF on the European Commission´s handling of a request for access to documents related to the launch of the EU Internet Forum (5 July 2017) paras 15–16.
[454] EDRi, 'EDRi and Access Now withdraw from the EU Commission IT Forum discussions' (31 May 2016) <https://bit.ly/2EJ7N3m> accessed 23 July 2020.
[455] ibid 5.
[456] See Chapter 3.2.1.
[457] Code of conduct, 1.
[458] 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 and national laws transposing it [2008] OJ L328/55.

best practices with other internet companies, platforms and social media operators'.[459] Thus, it is unsurprising that the general obligation is in line with free expression management: 'Companies, taking the lead on countering the spread of illegal hate speech online, have agreed with the European Commission on a code of conduct setting the following public commitments'.[460]

The general statement is followed by the more specific obligations that prescribe a suite of standards for removing hate speech that conveniently resemble companies' hard content moderation processes. The importance of efficiency is stressed throughout. For instance, the companies are to have in place 'clear and effective processes to review notifications regarding illegal hate speech' to remove such content. In addition, the companies need to 'provide information on the procedures for submitting notices, with a view to improving the speed and effectiveness of communication between the Member State authorities and the IT Companies'.[461] Vaguely enough, the notifications are to be managed by 'dedicated teams', which conveniently allows further efficiency to be gained by double outsourcing the work to sub-contractors.[462] However, the companies are to train their personnel on 'current societal developments', which arguably could require them to pay attention to the doctrinal evolution of fundamental rights as well.[463] After all, the companies are required to assess the notifications not only against their community standards but also law, 'where necessary'.[464]

In sum, the Code is deeply entrenched in managerial cooperation, deploying outsourcing in earnest. However, it is entirely a matter of faith whether and how the companies *de facto* adhere to their prideful free expression statements, or to hate speech removal either. To be sure, the Code prescribes public monitoring in the form of ongoing performance reporting. Monitoring of hate speech management is carried out by the Commission itself with 'a network of civil society organisations' tasked to aggregate the relevant quantifications.[465] The Commission has since the release of the Code published yearly reporting factsheets that include 'key figures' to measure success. The latest factsheet is from February 2019 including also the results of earlier monitoring rounds.[466] Apparently, this was also the year when the Code could be announced to have delivered a definitive success. Věra Jourová, then Commissioner for Justice, Consumers and Gender Equality, solemnly declared: 'Today, after two and a half years, we can say that we found the right approach [to hate speech] and established a standard throughout Europe on how to tackle this serious issue, while fully protecting freedom of speech'.[467]

---

[459] Code of conduct, 2.
[460] Ibid.
[461] Ibid.
[462] Ibid.
[463] Ibid.
[464] Ibid.
[465] MEMO/19/806.
[466] Commissioner for Justice, Consumers and Gender Equality and Directorate-General for Justice and Consumers, 'Code of Conduct on countering illegal hate speech online: Fourth evaluation confirms self-regulation works' (Factsheet 4th round) (February 2019) <https://bit.ly/3gmNlTD> accessed 26 July 2020.
[467] Commissioner for Justice, Consumers and Gender Equality, 'Countering illegal hate speech online – EU Code of Conduct ensures swift response' IP/19/805. See also the additional factsheet ticking all the standards met: Commissioner for Justice, Consumers and Gender Equality and Directorate-General for Justice and Consumers, 'How the Code of Conduct helped countering illegal hate speech online' (February 2019) <https://bit.ly/33fgsVv> accessed 26 July 2020.

The monitoring factsheet indicates that the best indicator of success is that companies have continuously increased speed, with most of the notifications assessed promptly within 24 hours, reaching an average of 89%.[468]

It seems that companies have also magically found almost a perfect balance between illegal hate speech and protected speech, which practically means that, in average, 72% of the illegal hate speech notified is removed. The Commission estimates this 'a satisfactory removal rate, as some of the content flagged by users could relate to content that is not illegal. In order to protect freedom of speech only content deemed illegal should be removed'.[469] Unfortunately, it is not disclosed how the Commission knows that 72% indicates an accurate fair balance. Removal rates, however, vary depending on the severity of hateful content. 85.5% of 'content calling for murder or violence against specific groups' was removed, while presumably the more controversial 'content using defamatory words or pictures to name certain groups' was removed only in 58.5 % of the cases. Purportedly, this proves that 'the reviewers assess the content scrupulously and with full regard to protected speech'.[470] One can but admire the confidence given that the company-specific removal rates indicated significant variation. For instance, YouTube removed 85.4 % of notified content, while Twitter only 43,5 %. The differences might reflect the respective tastes of the companies as, for instance, Twitter used to be known for its unbridled adherence to the First Amendment.[471] Be as it may, the differences are inscrutable, unchallengeable, and suggest rather arbitrary interpretation of the relevant fundamental rights concepts.

In September 2019, the Commission also published an extra report addressed to the Council, which assessed the company measures against some of the specific Code commitments. Again, the results are noted to reveal an overall success in terms of efficiency.[472] It remarks:

> The current average removal rate [of stable 70%] can be considered as satisfactory in an area such as hate speech, given that the line against speech that is protected by the right to freedom of expression is not always easy to draw and is highly dependent on the context in which the content was placed.[473]

As regards staff training, the standard is met as all companies conduct 'regular and frequent trainings, and provide coaching and support for their teams of content reviewers, including on the specificities of hate speech content'. Facebook is reported to hold discussions in every two weeks.[474] According to Kaye, within the company these discussions are referred to as 'mini-legislative sessions'.[475] Their agendas and minutes are

[468] Factsheet 4th Round (n 446) 1.
[469] ibid.
[470] ibid 3.
[471] See Josh Halliday, 'Twitter's Tony Wang: 'We are the free speech wing of the free speech party'' *The Guardian* (22 March 2012) <https://bit.ly/3gnKf1I> accessed 22 May 2020. The stance has, at least judging by official statements, changed a bit more balanced in the course of years. Olivia Solon, 'Twitter launches another bid to tackle bots and abuse after years of promises' *The Guardian* (1 March 2018) <https://bit.ly/30jtzCY> accessed 22 May 2020.
[472] Council of the European Union, 'Assessment of the Code of Conduct on Hate Speech online: State of Play' (27 September 2019) 12522/19, 2.
[473] ibid 3.
[474] ibid 4.
[475] Kaye (n 142) 53–58.

publicized.[476] By contrast, the sessions themselves are secret. By juxtaposing mini-legislative sessions with a democratic legislature, it becomes clear that only one of them is accountable to public.

What the monitoring reports do not explicitly mention is that for the EU, the Code has been highly *cost*-effective. Most costs ensuing from the commitments were, to large extent, already internalized by the companies. The oversight costs incurred from yearly factsheets, with civil society gathering information, can be expected to be modest. This keeps the cost-benefit calculus extremely positive. And to be fair, it is likely that the Code indeed helped to curtail some of the rampant hate on platforms. It would be a triumph if only utilitarian efficiency and well-meaning outcomes were the sole value at stake. The Code was also later complemented by a 'voluntary' Communication and a Recommendation.[477] In turn, the inscrutable EU Internet Forum carried on with solving the problem terrorism. In particular, it has since specialized in facilitating the transfer of 'best practices', especially filters and other algorithmic technologies.[478]

4.2 Directive on Audiovisual Media Services

The second case study concerns 'video-sharing platform services'. Commission announced its Proposal for amending the Directive 2010/13/EU (herafter in this sub-chapter the Proposal) in May 2016.[479] Directive (EU) 2018/1808 concerning the provision of audiovisual media services (hereafter the Directive)[480] came into force on 20 November 2018.[481] During the writing of this thesis, the period for implementation is still running with deadline being 19 September 2020.[482] While the Directive has been amended previously, this was the first time when these video-sharing platform services were included in its scope. Effectively, the most prominent service to be regulated is Google's YouTube video-sharing platform, but the provisions concern also social networking platforms like Facebook and Twitter to the extent these platforms host audiovisual content, that is, video.[483] I should note that the regulation of television broadcast and on-demand audiovisual services was amended as well. These amendments, however, are not the subject of study here.

[476] Facebook, 'Product Policy Forum Minutes' *Facebook Inc* (15 November 2018) <https://bit.ly/2EFDnPn> accessed 22 May 2020.
[477] Quintel and Ullrich (n 323) 182.
[478] European Commission Press Release, 'Fighting Terrorism Online: Internet Forum pushes for automatic detection of terrorist propaganda' IP/17/5105.
[479] European Commission, 'Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities' COM (2016) 287 final.
[480] Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities [2018] OJ L303/69.
[481] European Commission, 'Revision of the Audiovisual Media Services Directive (AVMSD)' (5 December 2018) <https://bit.ly/3194FoR> accessed 6 June 2020.
[482] PwC Ireland, '7 things you need to know about the proposed Online Safety and Media Regulation Bill' (29 January 2020) <https://pwc.to/2PeVu0N> accessed 6 June 2020.
[483] European Commission, 'Digital Single Market: updated audiovisual rules' MEMO/18/4093.

The rationale behind the changes was to ensure a 'level playing field' for the providers of different types of services in the increasingly converged media markets.[484] However, arguably the more important objective behind the inclusion of platform video content was the need to strengthen the protection of minors against harmful video content and protection against video content that incites violence or discriminatory hatred on the platforms.[485] Lastly, the Council included the aim to protect also against video which incites terrorism and other crimes, which made its way to the final text as well.[486]

4.2.1 Violence, Hatred, and Organizational Responsibility

During preparation, the Commission introduced something that, in my view, is a notable innovation in social media regulation: namely the concept of 'organizational responsibility'. The Impact Assessment notes: 'Video-sharing platforms employ tools like Autoplay (switched on by default for all videos in Youtube) which enable direct exposure to potentially harmful content and incitement to hatred'.[487] Personalization-based autoplay features opt for the video that will be played next.[488] The role of these 'tools' in facilitating harmful content is also noted with an illuminating example:

> In 2015, the video of two US journalists being murdered during a live broadcast spread quickly. When the video was taken down after 10/15 minutes, it had already been shared 500 times on Facebook. Due to the Autoplay feature, many users saw the video unwillingly in their news feed. Since the feature debuted on Twitter in June 2015, many people reported that it auto-played all videos, including exceptionally violent ones.[489]

Therefore, it is implied that organizational responsibility could precisely target the algorithmic personalization mechanisms that amplify the effects of violent or otherwise harmful content. In terms of adverts, as we saw in Chapter 2.3, the extremely innovative and lucrative advertising tools allow potentially harmful advertisements to be targeted precisely to the most amenable audiences. Proposal correctly disentangled harmful content from its facilitation by noting that:

> The system would be compatible with the liability exemption for hosting service providers set out in Article 14 [of the e-Commerce Directive], in as far as that provision applies in a particular case,

[484] Andrej Savin, 'Regulating Internet Platforms in the EU: The Emergence of the 'Level Playing Field'' (2018) 34 Computer Law & Security Review 1215, 1223–1225.
[485] COM (2016) 287 final, Explanatory memorandum, 13.
[486] Council of the European Union, 'Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities (First reading) – general approach' (24 May 2017) 9691/17, 45.
[487] European Commission, 'Impact Assessment Accompanying the document Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities' SWD (2016) 168 final, 5.
[488] See Chapter 2.3.
[489] SWD (2016) 168 final, 5 (paraphrasing Robinson Meyer, 'When a snuff film becomes unavoidable' *The Atlantic* (26 August 2015) <https://bit.ly/3fm5D6n> accessed 6 June 2020).

because these obligations relate to the responsibilities of the provider in the organizational sphere and do not entail liability for any illegal information stored on the platforms as such.

While regulating amplification of course affects freedom of expression as well, it may better address the undeniably thorny issue of how to deal with the proliferating of content which is harmful but may nonetheless be legal. The insights of organizational responsibility made their way to the final text in recitals 47 and 48.

Thus, the central provision concerning video-sharing platform companies is new Article 28b. It requires Member States to ensure that video-sharing platform providers take 'appropriate measures' to: protect minors from paid or user-generated video content which 'may impair their physical, mental or moral development'; protect the public from paid or user-generated video content 'containing incitement to violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the Charter'; and protect the public from paid or user-generated video content that constitutes a criminal offence in relation to terrorism, child pornography or racism and xenophobia. A non-exhaustive list of appropriate measures is included in Article 28b(3). However, while the ten-point list exhibits genuine effort to carefully define the actions to be taken, they reflect the idea of organizational responsibility, as outlined above, only thinly. For instance, under Article 28b(3)(d) the companies could be required to establish and operate 'transparent and user-friendly mechanisms for users of a video-sharing platform to report or flag to the video-sharing platform provider concerned the content referred to in paragraph 1 provided on its platform'. As a free expression safeguard, however, all *ex ante* monitoring including upload filtering was explicitly excluded. Point i) describes 'establishing and operating transparent, easy-to-use and effective procedures for the handling and resolution of users' complaints to the video-sharing platform provider in relation to the implementation of the measures referred to in points (d) to (h)', something that could be understood as another (company-enforced) free expression safeguard. However, paragraph 7 foresees a possibility for escalation to an impartial out-of-court redress mechanism, while redress in court remaining available under paragraph 8 as well.

To be sure, the measures do concern the 'organisational sphere'. But none foresees interference to the algorithmic personalization mechanisms. For instance, it is not stated that Member States could mandate that algorithmic features such as YouTube Autoplay would be set off as default, or that some harm-amplifying personalization features could be outright banned. However, a notable and positive point is that Member States may impose stricter or more detailed measures, as explicitly mentioned in Article 28(b)(6).

4.2.2 (Un)viability of Co-regulation

The question on how far the positive obligations are shifted within private companies will be highly contingent on what the 'appropriate measures' will eventually entail and how the impositions are enforced. In this regard, it seems the Directive informs a sort of mix of the two outsourcing techniques I described in Chapter 3.2, as Article 28b(4) encourages, though not strictly necessitate, the use of co-regulation in implementing the rules. The choice of co-regulation is justified only in recital 49: 'It is appropriate to involve video-sharing platform providers *as much as possible* when implementing the appropriate measures to be taken pursuant to Directive

2010/13/EU. Co-regulation should therefore be encouraged'.[490] The relationship with fundamental rights and co-regulation was not problematized during the legislative process. Therefore, the principle of non-applicability of co-regulation in fundamental rights issues seems indeed to bend in the Directive.[491] Impact Assessment underlined 'it is possible to identify scope for cost-efficiency also when looking not only at legal obligations but also at the possibility to leverage self and co-regulation arrangements'.[492]

Besides that co-regulation is plainly cheaper, according to the views collected by the Parliament Research Service at the start of the legislative process, it has two claimed benefits in media regulation. As co-regulation is supposed to be more adaptive to changes, it does not hinder innovation as much as rigid state regulation.[493] During the public consultation this was especially underlined by the ICT industry. For instance, EDiMA, which is perhaps the most prominent lobbying association representing 'online platforms and other innovative tech companies' in Europe,[494] stated that if any regulation is needed at all, '[t]he status quo would be complemented with self-/co-regulatory measures and other actions (media literacy, awareness-raising)'.[495] It stressed that during the revision process the EU should focus on two factors:

> To what degree the current AVMSD (and its articulation with the e-commerce Directive) has allowed digital innovation to drive the growth of the European creative sector and to empower consumers? To what degree extending the AVMS framework to other areas would harm the Internet, digital innovation and in turn the European creative sector?[496]

According to EDiMA, on platforms traditional media houses and individual cultural workers can engage in genuine competition for audiences. This fosters innovation.[497] Yet in reality, the competition for platform visibility is far from meritocratic but orchestrated by soft content moderation and pervaded by political-economic interests of anyone that is willing to pay for extra visibility. In addition, we should remember that innovation is not a virtue. The autoplay features that algorithmically opt and play videos from the endless repertoire of content have likely been breakthrough innovations. Even so, the Commission had now identified these same features as problematic examples of hate amplification.

The other justification for preferring self- and co-regulation is the empowerment of media users instead of state 'paternalism'.[498] Indeed, EDiMA's position hailed the liberating potential of the video-sharing and social media platforms. It declares that '[d]iscoverability of content has never been so easy, and put to such use. It

[490] Emphasis added.
[491] Cappello (n 295) 17.
[492] SWD (2016) 168 final, 105.
[493] European Parliamentary Research Service, 'The Audiovisual Media Services Directive: State of Play' (November 2015) PE 571.329, 11. Marsden notes the same justification, Marsden (n 299) 48.
[494] See EDiMA Website, <https://edima-eu.org/> accessed 8 June 2020. EDiMA includes all the biggest social media companies as its members.
[495] EDiMA, 'Consultation on Directive 2010/13/EU on audio-visual media services (AVMSD) A media framework for the 21st century: Questionnaire' (July–September 2015) 17 <https://bit.ly/33kKzuK> accessed 8 June 2020.
[496] ibid 7.
[497] ibid 8–9.
[498] European Parliamentary Research Service (n 493) 8.

has never been easier for AVMS providers to make their content discoverable and there exists a plethora of new tools, consumer personalisation technology and social recommendations'.[499] However, as we saw in Chapter 2.3 and 3.5.2, this paints a partial picture, as user choices and other online behavior feed into platform's algorithmic workings, nudging users and making it impossible to disentangle the effects that their choices have on information organization from the effects of other factors.

As we explored in Chapter 3, co-regulation that can strike some sort of balance between fundamental rights and flexibility/efficiency requires a robust public monitoring and enforcement system. In this regard, Article 4a(1)(c–d) of the Directive prescribes somewhat reassuringly 'regular, transparent and independent monitoring and evaluation of the achievement of the objectives aimed at' and 'effective enforcement including effective and proportionate sanctions'. But as I argued, robust enforcement adds to the transaction costs which strain utilitarian cost-benefit calculus. Indeed, it is expected that the co-regulatory oversight of video-sharing platforms under the new Directive will be a significant burden for public regulators.[500]

In addition, the level of opaqueness of the media regulator also affects how undemocratic a co-regulatory model ends up being. Independence of media and its governance have always been considered important for freedom of expression, but now the Directive further reinforced the independence requirements of Member State regulators in Article 30. Arguably, developments in certain Member States, Hungary's controversial Media Act adopted in 2011 as an example, might have played some role here.[501] Even so, inscrutable private control over media is not necessarily better or more democratic.

Article 28a(1) also reinforces the so-called 'country of origin principle' which determines the jurisdiction between Member States. For most video-sharing service providers, this practically means that they fall under the jurisdiction of Ireland where they established their main establishment in Europe.[502] Therefore, 'it could be said that preparatory work on the implementation of Article 28b of the Directive [on appropriate measures] appears almost redundant outside Ireland'.[503] This is why it's reasonable have a look into the implementation plans even though the scope of this thesis is generally delineated to remain on the European level. As regards audiovisual content on social media platforms in the whole Europe, the country of origin principle effectively shifts the site for cooperation and contest to one Member State.[504] There may be good justifications for reinforcing the principle. Yet in my view, it places excessive responsibility over the implementation of regulatory and administrative framework to a single regulator. Of course, concentrating regulation enforcement

[499] EDiMA (n 495) 26–27.
[500] Cappello (n 295) 21.
[501] European Parliamentary Research Service (n 493) 8.
[502] Cappello (n 295) 32–33.
[503] ibid 32.
[504] According to a Commission study, already in 2016 Facebook and YouTube together 'account for an estimated 72.2% of the monthly traffic to OSM [online social media] across the 28 EU Member States plus Norway and Iceland (43.6% Facebook and 28.6% YouTube), followed by Twitter (4.3%) and Instagram (2.5%)'. European Commission (n 13) 15. All the companies have their main European establishments in Ireland.

on the largest platforms to Ireland is nothing new. The GDPR enforcement against these same companies has ended up on the shoulders of the Irish Data Protection Commission as well.[505]

Again, the cards for Europeans' fundamental rights on platforms are set in Ireland's hand. Cappello notes that in practice, 'jurisdiction over the same overall service (content on a video-sharing platform) will sometimes be shared: Ireland will be responsible for the *platform* and the relevant member state for the *content played through it*'.[506] Also the new out-of-court complaint mechanisms will be regulated under Irish law.[507] In my view, this is problematic from the viewpoint of public accountability. The expressions should not be thought in isolation of the medium, i.e. platform itself, because algorithmic soft moderation sets the terms of for content's visibility. Moreover, the debates and harms on social media 'spill over' to affect the overall democratic deliberation in that society. The organization of content greatly influences the harm it may cause. In addition, jurisdictional concentration makes it harder for people to challenge the terms of algorithmic intermediation, impairing the horizontal effect of their rights.

4.2.3 Irish Online Safety & Media Regulation Bill

In early 2019, Richard Bruton, Minister for Communications in Ireland, made a keynote speech regarding the implementation of the Directive, where he declared that,

> The situation at present where online and social media companies are not subject to any oversight or regulation by the state for the content which is shared on their platforms is no longer sustainable. I believe that the era of self-regulation in this area is over and a new Online Safety Act is necessary.[508]

Thus, a '[r]egulator, an Online Safety Commissioner, would oversee the new system'.[509] Afterwards, a public consultation took place which sought to gather stakeholder views on how the specific co-regulatory framework prescribed in the Directive should be set up in Ireland.[510] Naturally, the major platform companies Google Ireland and Facebook Ireland put forward their views as well. Their contributions provide important insights of the specific challenges the Irish administrative framework will be facing in implementing co-regulation. Companies are invoking both the narratives of cooperation and contest. Facebook hails the achievements of

[505] After two years in force, it has become clear that the GDPR has so far somewhat comprehensively failed to curtail the extractive data protection practices of the largest companies, something that was arguably one of its initial objectives. Digital rights NGO AccessNow has recently labeled the GDPR enforcement 'an administrative crisis', pointing among others to the excessive pressure and burden piled on the Irish Data Protection Commission. AccessNow, Two Years Under the EU GDPR: An Implementation Progress Report (May 2020) 6, 12–14 <https://bit.ly/33jMCyU> accessed 15 June 2020.
[506] Cappello (n 295) 33. Emphasis added.
[507] ibid; and ERGA, 'Analysis & Discussion Paper to contribute to the consistent implementation of the revised Audiovisual Media Services (AVMS) Directive' (2018) 10 <https://erga-online.eu/?page_id=14> accessed 25 July 2020.
[508] Minister for Communications, Climate Action and Environment, cited in Department of Communications, Climate Action and Environment, 'Regulation of Harmful Online Content and the Implementation of the revised Audiovisual Media Services Directive' (April 2019) <https://bit.ly/3grnZEk> accessed 10 June 2020.
[509] ibid.
[510] ibid.

its hard content moderation apparatus and underlines its long, 15 years of experience on addressing harmful content.[511] It heavily underlines the effectiveness and efficiency of cooperation. For instance,

> The future regulator should also implement a flexible and responsive rapport with VSPS, developing a working relationship that jointly seeks to protect internet users from harmful content while also ensuring that the ability to innovate is preserved and service providers are able to adopt the best solutions to these issues.[512]

However, it is not considered that perhaps the best solutions are not *ex post* hard content moderation measures, but ones that target the root causes that make the harm to proliferate in the first place. This would require scrutinizing the platform code and algorithms instead of focusing solely on 'clear definitions on illegal content' as Facebook suggests.[513] Unsurprisingly, the company stays silent on its platform mechanisms. Nor it ponders whether it would be the job of the state instead of Facebook to decide the 'best solutions' to fulfill positive obligations.

In turn, Google Ireland took a more challenging stance. While it similarly self-congratulates its hard content moderation, it also implies that effectively any regulatory interference that goes beyond the notice-and-takedown model of the e-Commerce Directive could implicate freedom of expression and hamper innovation. To be sure, Google states the general preference for cooperation and even suggests a new multistakeholder forum to be set up for cooperative monitoring.[514] But on the one hand, the main insistence is that,

> In particular, we would highlight the need for national legislation to reflect the 'notice and takedown' procedures specifically envisaged by the E-Commerce Directive (…); this system strikes a careful balance between the interests of persons affected by unlawful information, ISSPs and internet users.[515]

Conveniently for Google, the e-Commerce Directive is totally blind for soft moderation by regarding platforms as completely neutral hosts. Google also demands transparency from *the government regulator* to guarantee that its intrusions are proportionate.[516] In sum, 'governments should take a flexible, collaborative approach that supports best practices, and promotes research and innovation' with the main role of the regulator again limited to providing clear definitions for unlawful content.[517] Taken together, all these suggestions denote a comprehensive obfuscation of the Directive mandates to water down organizational responsibility. This underlines how this concept, like a systemic duty of care, allows myriad interpretations depending on the

[511] Facebook Ireland Limited, 'Public Consultation on the Regulation of Harmful Content on Online Platforms and the Implementation of the revised Audiovisual Media Services Directive: Submission of Facebook Ireland Limited' (2019) 1 <https://bit.ly/2PirOjb> accessed 10 June 2020.
[512] ibid 9.
[513] ibid 6–7.
[514] Google, 'Public consultation on the regulation of harmful content on online platforms and the implementation of the revised Audiovisual Media Services Directive' (2019) 13 (Question 6) <https://bit.ly/2PirOjb> accessed 10 June 2020.
[515] ibid 1–2.
[516] ibid 7.
[517] ibid 6.

interests at stake. The company views are entirely in line with free expression management that seeks expressive and innovative immunities against interference to platform mechanisms.

On 10 January 2020, Draft Scheme of New Online Safety Law was publicized. According to the Minister: 'This new law is one of the first of its kind in the world and is breaking new ground in terms of how online services will be required to deal with harmful content'.[518] It would foresee a new Online Safety Commissioner under the new Media Commission as the regulator who would draft binding 'online safety codes' and oversee their enforcement. The appropriate measures would eventually be prescribed in the codes and would address the types of harmful content prescribed in the AVMS Directive.[519]

Online safety codes and even more precise 'guidance materials' would be used to provide the specific definitions of content types for platform companies, taking into account users' fundamental rights.[520] According to Head 50A of the draft scheme, the online safety codes are drafted by the regulator, with consultations of stakeholders, but not in a multistakeholder forum. They may prescribe measures regarding 'content delivery and content moderation'.[521] In addition, the Commissioner may appoint authorized officers to audit the companies. Under Head 15B, these officers would be vested with rather sweeping investigative powers resembling those of competition authorities.[522] The regulator could 'audit any user complaints and issues handling systems operated by designated online services and to direct a designated online service to take specified actions, including to remove or restore individual pieces of content and to make changes to the operation of their systems'.[523] As regards sanctions, while the early comments highlighted their stringency,[524] in reality, the big enforcement toolbox seems to prefer informal 'compliance and warning notices' and bargaining-style 'voluntary arrangement' negotiations over more coercive financial sanctions and blocking of service in Ireland. At least some of the negotiations, however, are stated to be public.[525]

Lastly, the enforcement includes the power to devise and operate so-called 'super complaints'. This is where 'nominated bodies', for instance NGOs and other Member State regulators could bring systemic issues they have identified on a platform to the attention of the Irish regulator for possible action.[526] The super complaint

[518] Minister for Communications, Climate Action and Environment, cited in Department of Communications, Climate Action and Environment, 'Minister Bruton Publishes Draft Scheme of New Online Safety Law' (10 January 2019) <https://bit.ly/3hZ7505> accessed 10 June 2020.
[519] Department of Communications, Climate Action and Environment, 'General Scheme of the Online Safety & Media Regulation Bill 2019: Amending the Broadcasting Act 2009' (hereafter in footnotes General Scheme) (January 2019) 81–83, Head 50A <https://bit.ly/33jN0xq> accessed 10 June 2020.
[520] ibid 85–86, Head 51.
[521] ibid 81–83.
[522] ibid 36–40, Head 15A and Head 15B.
[523] ibid 87.
[524] See eg, Colm Gorey, 'Online safety law could block social media sites for non-compliance' *Siliconrepublic* (10 Jan 2020) <www.siliconrepublic.com/companies/online-safety-law-ireland-details> accessed 10 June 2020; and 'Plans for 'first-of-its-kind' online safety regulator published' *Irish Legal News* (10 January 2020) <https://bit.ly/3k5A0kY> accessed 10 June 2020.
[525] General Scheme, 88–93, Head 53, Head 54 and Head 55.
[526] ibid 88.

mechanism is meant to facilitate the cooperation between Member State regulators foreseen in the Directive and could thus offer a channel for other Member States to influence the regulation of the biggest platforms.

On paper, it may be that the Irish draft scheme takes positive obligations seriously by relying on binding and robust regulation and bringing in much-needed *public* responsibility for the protection of rights. However, much remains to be seen. A compromise would be contingent on the funding and actual regulatory practices of the Commissioner regulator and at this point one cannot rule out the occurrence of an administrative crisis similar to one troubling the GDPR enforcement. The lingering issues on how to understand and regulate complex code and algorithms effectively are hard to tackle. In addition, as an access point to company contest, under Article 28b(3) all imposed measures 'shall be practicable and proportionate'. According to Cappello, the requirement practically means that company challenge against too incursive regulatory efforts is likely.[527] One could expect the fiercest contest to be mounted against measures that target automated platform mechanisms. Indeed, Google had already reminded that companies should be able to challenge all measures in court, possibly resulting in years-long battles that are indeed cost-intensive for regulators. In line with free expression management, '[c]ompanies should be able to bring evidence before a neutral arbiter to contest findings of systemic failure [of obligations] and raise countervailing considerations, such as respect for international human rights'.[528] Again, the alternative narrative here is that the obligations may need to be imposed *precisely because* the state has to fulfill its human rights obligations.

Even so, at least the draft administrative framework has the potential to subject both hard and soft moderation to some public control, even though the efficiency/democracy trade-off is of course part of the package. As noted in Chapter 3.3 and 3.4, the oversight mechanisms may themselves be opaque and unaccountable to the public, in this case especially to other European citizens which are lacking regulatory capabilities to influence regulation platforms which, of course, affect them too. Yet, the framework with broad administrative powers may still serve as a site for innovative development of another kind of 'best practices', that is, novel ways for more democratic fundamental rights protection against platform power and business model.

4.3 Regulation on Preventing Terrorist Content Online

The third and most recent case study is the Proposal of a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online (hereafter terrorist content regulation or in this sub-chapter, the Proposal).[529] The Proposal aims to prevent the dissemination of terrorist content on the platforms.[530] The Commission announced the proposal on 12 September 2018 and the Council and the European Parliament both proposed amendments to it on 3 December 2018 and 17 April 2019 respectively,

[527] Cappello (n 295) 21.
[528] Google (n 515) 7.
[529] European Commission, 'Proposal of a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' COM (2018) 640 final.
[530] COM (2018) 640 final, Explanatory memorandum, 3–4, 7.

with the Parliament's amendments constituting major revisions to the original text.[531] At the time of writing this thesis, the legislative process is still under way at the closed trilogue negotiations phase.[532] Since the final text of the regulation is still debated, my case study focuses on the initial Commission proposal but also analyzes the relevant amendments made by the Council and the Parliament.

4.3.1 EU's Attempt of Fiat Interdiction

Of all the three cases examined in this chapter, terrorist content regulation follows the logics of fiat interdiction to the furthest. Already the choice of instrument, an EU regulation that is directly applicable without Member State implementation, to some extent reflects high priority. While the Commission states a regulation to be the most appropriate instrument for the functioning Digital Single Market, clarity, and legal certainty,[533] it also justifies the choice by the extraordinary threat that terrorist content creates for security in Europe. Extraordinary threats cannot be tackled by soft means only. Thus, the explanatory memorandum, after lauding the EU Internet Forum, takes note of the 'limitations' of cooperation.[534] In effect, for the Commission, these limitations mean friction in the form of insufficient amount of removed content. The impact assessment admits the underlying logic of fiat interdiction more candidly. It acknowledges that the cooperative strategy has delivered promising results in removing terrorist content.[535] However:

> Taking into account *the particular need and urgency* to prevent the dissemination of terrorist content online, there is a corresponding need to put in place measures which capture the most harmful content and are particularly effective.[536]

In addition, 'the limited level of progress when compared to the scale of the problem, developments and the need to significantly reduce accessibility to terrorist content as a matter of urgency reveal the limits of the voluntary approach'.[537] To achieve full compliance by big and small platform companies alike, all the retained

[531] Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online – general approach' (3 December 2018) 14978/18; and European Parliament legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, P8_TA (2019) 0421.
[532] The negotiations started with considerable differences in positions between the Parliament and the other two bodies. The process is further hindered by the ongoing COVID-19 pandemic which led to the suspension of negotiations in spring 2020. Jens-Henrik Jeppesen, David Nosák and Pasquale Esposito, 'EU Tech Policy Brief: March 2020 Recap' *Center for Democracy & Technology* (3 April 2020) <https://bit.ly/3gyq3KA> accessed 4 May 2020.
[533] COM (2018) 640 final, Explanatory memorandum, 3–5.
[534] ibid 1–2.
[535] European Commission, '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' SWD (2018) 408 final, 4.
[536] SWD (2018) 408 final, 48. Emphasis added.
[537] ibid 24. In terms of urgency, the assessment was revised in the process since, pointedly, the Regulatory Scrutiny Board had noted that '[t]o justify the choice of scope, the report should explain why, despite numerous ongoing initiatives, there is a more urgent need to act now on terrorist content'. SWD (2018) 408 final, Annex I: Procedural Information, 56.

options in the proposal are legally binding.[538] The proposal also reminds of the 'particular serious consequences of certain terrorist content'.[539]

The perceived threat is ultimately rooted in the terrorist attacks occurred in Europe in the last two decades.[540] However, the Commission attributes the exceptional gravity also to the unprecedentedly fast spreading of terrorist content on platforms. Given the lament over virality, it's first seems strange that the Commission does not reiterate the notion of organizational responsibility. Yet the reason for rejecting direct interventions to platform mechanisms is the seductive efficiency of platform moderation. Too direct attempts against companies' business would risk fruitful cooperation in the future. Indeed, despite preference to unilateral interdictions, the Commission also acknowledges that cooperation with the industry is still needed. Impact assessment reminds that 'action by Member States alone to refer or block material, will always fall short considering it is the companies who are best placed to know which technology can be applied and how in order to ensure a safe online environment for their users'.[541]

The need for maximal efficiency explains the focus on terrorist content, not its facilitation. The Commission adopts the narrative of system-abuse and sets out how viciously terrorists exploit the platform mechanisms intended for facilitating free expression and information.[542] In turn, the companies play their victim role well. During the consultations, companies expressed how terrorist content hampers their business and damages reputation.[543] Therefore, like the previous case studies, the proposal ends up mixing both cooperation and contest.

4.3.2 Removal Orders by Competent Authorities

Fiat interdiction is also visible in the measures proposed. Firstly, the Proposal mandates each member state to designate a 'competent authority', which is given power to scour the platforms for potentially terrorist material and to assess whether this content falls within the definition of terrorist content.[544] Thus, the task of the competent public authority is essentially to determine the precarious line between the content that is protected by freedom of expression and the content that is terrorist and may then be justifiably taken down. Where the competent authority deems content terrorist, according to Article 4(1) of the proposal, it has the power to issue a removal order to the platform company, after which the company takes the content down. According to

[538] SWD (2018) 408 final, 25. It should be noted that, according to Article 2(1), the proposal applies to all 'hosting service providers' and not only the biggest social media platforms to which this thesis is largely limited. For the sake of consistency in terminology, in this sub-chapter my preferred term 'platform company' equates with 'hosting service provider'.
[539] Recital 30. It remains contestable whether the threat of terrorist content constitutes a more urgent threat to European societies than other issues of a platform setting. For instance, while 61% of respondents to the dedicated Eurobarometer claimed to have seen some type of illegal content, only 6% of respondents claimed to have seen content that was specifically terrorist. SWD (2018) 408 final, Annex 2: Stakeholder Consultations and Feedback, 62.
[540] COM (2018) 640 final, Explanatory memorandum, 1.
[541] SWD (2018) 408 final, 18.
[542] ibid 6; Proposal, Explanatory memorandum, 3.
[543] SWD (2018) 408 final, 6–7.
[544] COM (2018) 640 final, Explanatory memorandum, 10.

Article 4(2), the company must execute the order within one hour from the receipt of the order and this timeline may be extended only in narrowly defined exceptions stated in Article 4(7–8).

The introduction of removal orders was received with heavy criticism. Together with the wide definition of terrorist content laid down in Article 2(5), the orders were thought to lead to government overreach and over-removals. The one-hour timeline was seen unreasonable especially for smaller platform companies.[545] The criticism is not unfounded. As it becomes clear from the Impact Assessment, the aim of the Proposal was to provide the broadest possible mandate for the competent authorities so that potentially terrorist content could be removed as much as possible. Among of the policy options considered, the Commission chose the one with the broadest definition for terrorist content and the broadest obligations for removal, even though the same option was also acknowledged to be the most notorious for people's rights. The one-hour timeframe and narrowly defined exceptions of non-compliance seek to institute removal orders as a mechanism that forecloses any secondary assessment by the platform companies.[546] In line with fiat interdiction, the additional regulatory burden of such a unilateral enforcement is no issue for the Commission either, as the benefits for the vital cause of counter-terrorism would be so significant as well.[547]

However, despite draconian competences provided for authorities, the removal order mechanism is actually the most accountable part of the whole Proposal. The commentators welcomed the use of binding law for online regulation as well.[548] Indeed, the Commission noted that to challenge removal orders, judicial redress, as an important safeguard, would be available for the affected users and the platform companies alike.[549] Also the Impact Assessment recognized that, '[r]egarding removal orders, the main responsibility of assessing the content and ordering the removal stays with the competent authority, which would have to undertake a thorough check of the content and be subject to legal controls before the issuing of the order justifying the one hour removal time for this measure'.[550]

Against this backdrop, it is peculiar, that no mention of judicial redress has made its way in the actual articles of the Proposal but only in recital 8. Be as it may, the Parliament took note of this uncertainty by adding a new Article 9a on effective remedies in its own version. This does not, however, mention judicial redress specifically but mandates Member States only to 'put in place effective procedures for exercising this right'. Given the relatively low interest in an isolated case of removed content against the heaviness of full-blown judicial process, the vague terminology could allow fruitful exploration of middle ground between efficiency

[545] Eg, 'Letter of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism' (7 December 2018) OL OTH 71/2018, 6; and Aleksandra Kuczerawy, 'The Proposed Regulation on Preventing the Dissemination of Terrorist Content Online: Safeguards and Risks for Freedom of Expression' *Center for Democracy and Technology* (5 December 2018) 9 <https://tinyurl.com/y4mr7wyj> accessed 15 March 2020.
[546] COM (2018) 640 final, Explanatory Memorandum, 7; Impact Assessment, 42.
[547] SWD (2018) 408 final, 49.
[548] Eg, Special Rapporteurs (n 546) 2.
[549] COM (2018) 640 final, Explanatory Memorandum, 5, 7.
[550] SWD (2018) 408 final, 42.

and procedural fairness. As Kuczerawy summarizes: 'The effectiveness of the remedy is questionable if it takes one hour to remove content, but months or years to put it back online'.[551] Ideally, Member States could develop new *publicly controlled* redress mechanisms that could be efficient enough but provide also better impartiality and public accountability. In these redress mechanisms, users could at least in principle assert their rights horizontally against companies too. In this regard, it is somewhat discouraging, however, that during the public consultations, Member States wanted to express their special concerns that creating dispute resolution and oversight mechanisms would be burdensome and cost-intensive.[552]

In addition, the Parliament makes some other amendments to the Proposal that may be said to add some robustness to safeguards. They mitigate the most draconian aspects of removal order mechanism. Firstly, it narrowed down the definition of terrorist content. For instance, the Parliament's version institutes a blanket exclusion to content that is disseminated for 'educational, artistic, journalistic or research purposes' in new Article 2(2a). The Commission regretted this exception as 'creating a potential loophole' for terrorists.[553] Secondly, in new Article 2(9a), it redefined the competent authority to denote 'a single designated judicial authority or functionally independent administrative authority'. This seems to answer the calls for increased independency of the designated authority from law enforcement.[554] Thirdly, the Parliament introduced a separate Article 8a on certain transparency requirements for competent authorities, while the original Proposal introduced obligations only on platform companies. However, the one-hour time to comply with orders under Article 4(2) persisted. Even so, despite some lingering overreach concerns, from the perspective of this thesis, the Parliament's version seems to at least have the potential to tame the undemocratic excesses of fiat interdiction.

4.3.3 Referrals to Platform Companies

The second measure of a new mandated referral mechanism, in turn, is deeply entrenched in managerial cooperation. As mentioned, the Commission stressed the need to continue and intensify cooperation on the side of any unilateral power assertions. Consequently, Article 5(1–2) of the Proposal states that

> '[t]he competent authority or the relevant Union body may send a referral to a hosting service provider. Hosting service providers shall put in place operational and technical measures facilitating the expeditious assessment of content that has been sent by competent authorities and, where applicable, relevant Union bodies for their voluntary consideration'.

[551] Kuczerawy (n 546) 10.
[552] SWD (2018) 408 final, 29.
[553] European Commission, 'Follow up to the European Parliament legislative resolution on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' SP (2019) 440.
[554] FRA (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' (12 February 2019) FRA Opinion 2/2019, 7–8; Special Rapporteurs (n 546) 13.

Article 5(5) further stipulates the instruction for the company assessment: 'The hosting service provider shall, as a matter of priority, assess the content identified in the referral against its own terms and conditions and decide whether to remove that content or to disable access to it'. The model for the referral system comes from the Europol Internet Referral Unit (hereafter EU IRU). EU IRU was established in 2015 to search for potentially terrorist content on platforms and refer its findings to companies,[555] thus implying that the content should be taken down as a violation of community standards. It has been operating ever since. According to Vieth, the unit is part of 'the emerging constellation of public and private actors cooperating to regulate expression online in the name of counter-terrorism'.[556] Recital 15 of the Proposal clarifies that the relevant Union body issuing referrals would be Europol. Since 23 Member States do not have referral units of their own,[557] newly founded units would significantly contribute to the institutionalization of the model.

As the final decision upon removal of referred content would be made by the platform company, finding the fair balance between free expression and counter-terrorism would be transferred to platform companies. Moreover, the content subject to referrals is arguably going to be at the margins of legal if it does not fall under the more efficient removal order mechanism. Of course, officially the company is only interpreting its own community standards, not any binding fundamental rights instrument. But it is hard to deny this *de facto*, especially since the circumspect general 'duty of care' for platform companies in Article 3 stipulates that in complying with the regulation, companies shall ensure the fulfillment of the fundamental rights of the users and 'take into account the fundamental importance of the freedom of expression and information in an open and democratic society'.

So, who would be the companies' adjudicators, the real people making the decisions upon referrals? The Impact Assessment seems to anticipate extra costs for platform companies that would incur from the hired experts required for the more in-depth assessment of referrals.[558] However, the Proposal text does not impose any detailed requirements on the assessment of referrals, except that, according to recital 33, the handling process must be fast.[559] The Impact Assessment points out that '[s]ome of the costs [for the companies] are assumed to be absorbed in moderation functions already in place for other types of content'.[560] This conveniently leaves the door open for double outsourcing of referral assessment, transferring the positive obligations further down the regulatory chain.

---

[555] EU Internet Referral Unit, 'Transparency Report 2018' (20 December 2019) 3 <https://bit.ly/33i15LS> accessed 15 May 2020. EU IRU was founded only two months after the terrorist attacks in Paris against the staff of Charlie Hebdo magazine and the Jewish supermarket Hypercacher. Kilian Vieth, 'Policing 'Online Radicalization': The Framing of Europol's Internet Referral Unit' in Ben Wagner, Matthias C Kettemann and Kilian Vieth (eds), *Research Handbook on Human Rights and Digital Technology: Global Politics, Law and International Relations* (Edward Elgar Publishing 2019) 326–327.

[556] Vieth (n 555) 320.

[557] SWD (2018) 408 final, Annex 3: Who is affected and how?, 94

[558] ibid 94–95.

[559] The one-hour timeframe of the removal orders does not extend to referrals. COM (2018) 640 final, Explanatory memorandum, 5.

[560] SWD (2018) 408 final, 36–37.

The motivations of this state action laundering are twofold, as set out in Chapter 3.2. First, security policy could potentially be pushed beyond traditional constitutional limits. The Impact Assessment admits that referrals are not strictly limited to illegal content and they cannot be challenged in court as they are not legally binding.[561] Earlier, in the context of EU IRU the advocacy was more open. The EU Counter-Terrorism Coordinator pointed out to Member States that,

> Consideration should be given to a role for Europol in either flagging or facilitating the flagging of content which breaches the platforms' own terms and conditions. These often go further than national legislation and can therefore help to reduce the amount of radicalising material available online.[562]

Secondly, from the perspective of utilitarian cost-efficiency its's fully sensible to outsource. Removed content increases and, by contrast, the costs for Member States are expected to be none.[563] Not unrelated to savings in costs are savings in effort, as during the public consultations '[p]ublic authorities outlined the increasing difficulty to judge which content is harmful or illegal and which is not'.[564] Judged against this, it is rather unsurprising that the specific freedom of expression safeguards in relation referrals are outsourced to companies as well. The incumbent Article 10 is an emblematic act of outsourcing, including a highly open-ended mandate for the platform companies to devise 'effective and accessible mechanisms' for redress against referral-based removals. A lot will be left hanging from the vigilance (and resources) of the competent authorities who are tasked with general overseeing of complaint mechanisms. Moreover, the two-lane redress mechanism underlines the arbitrariness of public/private divide here. While from the perspective of user, it is irrelevant whether content is removed based on an order or a company decision after a referral, the user has the access to public review only where the content has been removed under a (legally binding) removal order.

The interconnection between outsourcing and free expression management becomes clear with the referral mechanism, as discretionary power granted for platform companies affords access points for their contest. In the case of EU IRU, the compliance ratio of referrals has varied between 89% in Q4 2017 to 63% in Q2 2018,[565] but on average the referred content has been removed in 86% of the cases.[566] This rate is intended to remain high and execution easy because of 'high quality' preliminary assessments made by the competent authorities.[567] Thus, there must be a difference between standards,[568] or at least some secondary assessment, to explain the gap. The power to decide allows companies to pick cases for challenge by refusing removal and then publicizing them as anecdotal examples of valiant opposition in their transparency reports. Aggregated

[561] SWD (2018) 408 final, Annex 5: Impact on Fundamental Rights, 103.
[562] EU Counter-Terrorism Coordinator, 'EU CTC input for the preparation of the informal meeting of Justice and Home Affairs Ministers in Riga on 29 January 2015' (17 January 2015) DS 1035/15, 3.
[563] SWD (2018) 408 final, Annex 3: Who is affected and how?, 95.
[564] SWD (2018) 408 final, 65.
[565] ibid 65.
[566] Europol, 'EU Internet Referral Unit - EU IRU: Monitoring terrorism online' (2020) <www.europol.europa.eu/about-europol/eu-internet-referal-unit-eu-iru> accessed 15 May 2020.
[567] SWD (2018) 408 final, Annex 9: Terrorist content online – evidence summary, 140.
[568] Vieth (n 555) 330.

figures of referral outcomes for monitoring are mandated in Article 8. The removal ratio of EU IRU alone should, however, reveal how much faith people could put in platform companies for their rights protection. According to the Commission, it has been between 96 and 100% for Facebook, YouTube, Microsoft and Twitter.[569] In effect, the referral mechanism is an example of bargaining enforcement. As Keller and Leerssen note, referrals are about 'convincing platforms to adopt law enforcement agents' interpretation'.[570] At the same time, '[b]oth assessment processes are obscure and completely inscrutable'.[571]

The Parliament removed the referral mechanism entirely in its own version, which the Commission noted with regret, as 'the deletion undermines the functioning of an effective cooperation tool between providers, the Member States and Europol'.[572] One could therefore assume the referrals to be one of the most contested part of the ongoing trilogue negotiations. The Commission receives backing from the Council, which seemed to be all in for broadening the referral system, making only minor adjustments to the mechanism in its own version.[573] It remains to be seen whether the Parliament holds to the end, however, as possible sign of compromise, it added a new recital 27a recognizing that '[r]eferrals by Europol constitute an effective and swift means of making hosting service providers aware of specific content on their services'. According to the recital, no changes are intended to be made to Europol's referral mandate under Regulation (EU) 2016/794.[574]

4.3.4 Companies' Proactive/Specific Measures

The third important measure for tacking dissemination of terrorism in the Proposal is the obligation to take proactive measures to protect platforms against terrorist content, as stipulated in Article 6. The competent authority may first request a platform company to take proactive measures, meaning actions that exceed mere reactionary measures taken after the company has been informed on the terrorist content. The measures should prevent the re-upload of content that had been already removed as terrorist, but also detect, identify and remove new content.[575] If the platform company does not cooperate or the competent authority deems the taken measures insufficient, according to Article 6(3) it may first issue another request for additional proactivity. It seems that as a baseline, the companies are again expected to innovate the measures themselves. If cooperation doesn't work, according to Article 6(4), the competent authority may impose a unilateral decision on the company, specifying the additional measures on behalf of the company. In the Council's version the competent authority's power to decide upon the nature and scope of measures under Article 6(4) was reinforced.[576]

[569] SWD (2018) 408 final, Annex 9: Terrorist content online – evidence summary, 135.
[570] Keller and Leerssen (n 168) 26.
[571] Vieth (n 555) 330.
[572] Commission follow up, SP (2019) 440.
[573] See generally, Council general approach 14978/18.
[574] Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA [2016] OJ L135/53.
[575] Article 6(2b). COM (2018) 640 final, Explanatory memorandum, 7.
[576] Council general approach 14978/18, 36.

Article 6(1–2) is probably meant to be read in conjunction with the systemic duty of care in Article 3, but what the Commission has especially in mind with proactive measures is couched in the technologically neutral term of 'automated tools'. In practice, these can be re-upload filters, upload filters, and algorithmic detection tools among others (and who knows what else in the future). In the Commission's vision, these could take the efficiency of content removals to a whole different level. The preparatory materials show how the efficiency of biggest companies' technology holds the Commission in thrall. Automation is unmatched in speed and the volumes of removed content immense. The Commission notes that 'some companies using automated detection claim that content can be identified within a matter of minutes or less'.[577] Experiences from the UK are also referenced. For instance, it is noted that tests have shown that a tool developed by a company named ASI Data Science can 'automatically detect 94% of Daesh propaganda with 99.995% accuracy'.[578] The conclusion is that, despite some lingering concerns, machine learning and artificial intelligence are becoming more and more promising.[579] Indeed, long regulatory chains are possible here. The Commission even envisions that a whole new market could be created for companies like ASI Data Science, who could develop solutions for platform companies to detect and filter terrorist content amongst legal content.[580]

Accuracy is crucial so that no lawful expressions are removed erroneously (false positives), and no illegal terrorist content is left on platforms (false negatives). All the high accuracy claims imply that some sort of sweet spot, an ideal fair balance is attainable and has almost been found between terrorist content and other lawful expressions. According to the Commission and the companies, it is becoming closer all the time.[581] But against whose interpretation is 94% of Daesh propaganda detected with 99.995% accuracy? It is plainly acknowledged that 'no data is available on the parameters under which a piece of content is considered to be terrorist material'.[582] At the development stage there are people interpreting, i.e. making sense of all these fluid concepts: Daesh propaganda, terrorist content (definition in Article 2 of the Proposal), journalistic speech and so on. How can people scrutinize and influence these underlying interpretations that can be at the end of the regulatory chain? Of course, there is no 100% to be reached as new situations in life force people, courts, authorities, legislators and others to continuous re-interpretation, leading to development that reformulates these concepts, including the reformulation of rights. How is this development in life considered and how the fair balance developed? The Commission notes indifferently that retraining of tools by someone is needed so that new patterns in terrorist content would be recognized. It stresses that efficiency must be maintained.[583]

To be sure, the Proposal includes safeguards here too. Firstly, under Article 3 the platform companies naturally have again a general obligation to guarantee users' freedom of expression, when they fulfill their obligations. This makes the tools definitively 'balancing technologies' I addressed in Chapter 3.3. The more specific

[577] SWD (2018) 408 final, 13.
[578] ibid, Available technologies for detection and takedown of illegal content, 144.
[579] ibid.
[580] ibid, Annex 3: Who is affected and how?, 93.
[581] ibid, Available technologies for detection and takedown of illegal content, 144.
[582] ibid 143.
[583] ibid 144.

safeguards are found in Article 8, 9 and 10. Under Article 9, the most important safeguard is 'human oversight and verifications', especially in cases where 'a detailed assessment of the relevant context is required in order to determine whether or not the content is to be considered terrorist content'. The provision conveniently allows further outsourcing to sub-contractors, as no further specifications is set for this human review requirement. Secondly, the importance of 'human in the loop' is underlined.[584] It is a final touch of human interpretation and empathy that supposedly can make the configuration imperfectly perfect. Unfortunately, there is little consideration on how 'human review' could be subjected to robust democratic oversight or how the so-called 'automation bias', humans' propensity for quasi-automated rubber-stamping of fully automated suggestions, could be countered.[585]

Article 15(1) and 17(1)(c) mandate Member States to coordinate and oversee the implementation of proactive by companies. Thus, in principle this could secure some control by the competent authorities, but in practice, the amount of oversight remains unclear. If a user's content is removed as a result of proactive measures, the user must seek redress from the online complaint mechanism managed by the platform company. Pointedly, it seems that the user is not explicitly entitled to challenge a proactive measure *in abstracto*, but only the removal of a piece of content. For the companies, however, under Article 6(5) they can indeed revoke the authority's request on proactive measures, where the authority needs to conduct a secondary review on its request. If it then issues a unilateral order, it is legally binding on the company and, thus, challengeable in court.

The Parliament considerably softened the obligations of Article 6. It renamed activities under Article 6 as 'specific measures' so that the terminology would appear less controversial against the prohibition of general monitoring obligation established in Article 15 of the e-Commerce Directive.[586] The deployment of automated tools was narrowed down with more precision on the technologies used, and the new specific measures are meant as voluntary. The new Article 6(1) stipulates only that the companies *may* adopt specific measures, and naturally, with full regard to freedom of expression. Thus, the amended article is not so much a mandate but rather a competence to be used or left unused by the companies. In practice, objectively it may be impossible to determine which measures have been requested by authorities and which are genuinely implemented by companies' own initiative. Yet little indication is given on what the measures would entail, although, recital 16 is added with a list that they 'may include regular reporting to the competent authorities, increase of human resources dealing with measures to protect the services against public dissemination of terrorist content, and exchange of best practices'.[587]

[584] ibid, Annex 5: Impact on fundamental rights, 105.
[585] Ben Wagner, 'Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems' (2019) 11 Policy & Internet 104, 117–118; and Council of Europe (Rapporteur Karen Yeung), 'Responsibility and AI: A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework' (September 2019) Study DGI (2019) 05, 64–65.
[586] See P8_TA (2019) 0421, 22, amended recital 19.
[587] Parliament position, 19. In addition, measures must be 'appropriate, targeted, effective and proportionate', effectively requiring careful contextual balancing of rights and interests.

Under the amended Article 6(4), competent authority can *only request* additional specific measures where the amount of removal orders indicates that the platform is not complying. However, the exact legal nature of such a 'request' is notoriously blurred. Its voluntary nature is ultimately cast into doubt by the fact that the amended Article 18(1)(d) foresees penalties applicable only to 'systematic and persistent breaches of the obligations' by platform companies 'following a request imposing additional specific proactive measures'.[588] Despite the apparent benignity of the term, it seems that requests are imbued with some coercive force. Nevertheless, the stated preference is clearly on cooperation. However, meaningful free expression protection would require scrupulous interrogation hash databases and algorithms, also when relevant technologies are sourced from third-party vendors. This is painstaking and possibly expensive.[589]

The transparency requirements under Article 8, with only slight modifications by the Parliament, impose ongoing reporting mandates of removal figures, amounts of appeal, general information on company actions and 'a meaningful explanation of the functioning of specific measures'. Bargaining enforcement that underlines cooperation in compliance is advocated throughout the original Proposal[590] and the Parliament only enhanced this.[591] As the EU-level monitoring on specific measures is to be facilitated by EU Internet Forum and coordination by Europol,[592] one can justifiably raise concerns whether the cooperative monitoring/enforcement will be subject to robust democratic scrutiny. To summarize, it seems that the Parliament was somewhat more successful in scrutinizing the more straightforward removal mechanism than the cooperative parts of the Proposal. This might inform a presumption by the Parliament that companies' governance is somehow more legitimate and harm-free than public. But such governance, as I highlighted in Chapter 3, may be even more undemocratic.

5. Conclusion: Between Co-optation and Compromise

We have seen that all three case studies exhibited the hybridity between public and private power, albeit in different varieties. The most prominent feature of the emerging legislative and administrative framework is the interdependency between platform companies and public regulators. Although, the proposal on terrorist content is clearly the most unilateral regulatory attempt of the EU, while the code of conduct on hate speech involves the companies to the furthest. Even so, the Code included a role for the Commission too. But in all case studies the role of public regulators is more ancillary. In turn, the companies are largely taking the responsibility for positive free expression and related obligations on social media, thus thoroughly merging

[588] The same is stated in the amended recital 19.
[589] In this regard, FRA called for general reinforcement of Member States' positive obligations by suggesting a specific recital that would 'refer to the positive obligation of the Member States to secure the effective exercise of fundamental rights and prevent fundamental rights violations, (…) and *underline that the effectiveness of software used to detect terrorist content should be adequately tested, especially from a fundamental rights perspective*. FRA (n 554) 11, emphasis added. This might better secure the extension of public fundamental rights scrutiny to the very end of the outsourcing chain. However, such a recital was not added.
[590] Eg, COM (2018) 640 final, Explanatory memorandum, 7; and SWD (2018) 408 final, 22.
[591] P8_TA (2019) 0421, recital 1(b), recital 27(a), article 13(4a).
[592] SWD (2018) 408 final, 50, 140–141.

private and public interests and correspondingly collapsing the border of between public and private, as foreseen by Black in her theory of decentred regulation and observed in rights context by Laidlaw.

In this chapter, I first reflect the emerging framework against the ECtHR jurisprudence. Secondly, I draw together the arguments and reflect on the effects the framework may have for freedom of expression. The final sub-chapter is more exploratory as it contains a brief reconstruction exercise by considering what structural changes should be regarded to maintain freedom of expression as an avenue for political equality.

5.1 Regulatory Framework and the ECtHR Jurisprudence

So, the roles of public regulators and companies in the legislative and administrative framework are changing through the interconnected ways of cooperation and contest. We must now see if we can determine whether the emerging regulatory framework is an 'adequate' one under the jurisprudence of the ECtHR. After all, according to the doctrine, the locus of positive obligations is on the state instead of private actors.

Firstly, it should be noted that the ECtHR jurisprudence that concerns user-generated content does not clearly address the issue of private involvement in regulation. For instance, in the most notable case *Delfi*, the ECtHR didn't scrutinize the privatization of platform monitoring.[593] The case was only about the *liability* of platform over content.[594] It was not considered whether in some cases the monitoring measures of service providers would actually require something more from the state's part than just *ex post* imposition of damages. Similarly, other notable cases, including *MTE v Hungary*, revolved around the question of service provider's liability over potentially illegal content.[595] In those cases, there was no more interference needed from the state's part.[596]

Moreover, in line with its generally positive stance toward the relationship between the Internet and free expression,[597] the ECtHR has tended to underline the positive impact of platforms for free expression. For instance, both YouTube[598] and Instagram[599] have received praise from the ECtHR as important means of exercising the freedom to receive and impart information. It has even accepted that 'YouTube is a unique platform on account of its characteristics, its accessibility and above all its potential impact, and that no alternatives were available to the applicants'.[600] While in *Delfi* it was accepted that the novel threats of the

[593] Marta Maroni, 'The Liability of Internet Intermediaries and the European Court of Human Rights' in Tuomas Ojanen and Bilyana Petkova (eds), *Fundamental Rights Protection Online: The Future Regulation Of Intermediaries* (forthcoming, Edward Elgar Publishing 2020) 239–240 (in manuscript).
[594] *Delfi v Estonia* ECHR 2015–II 319.
[595] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v Hungary* App no 22947/13 (ECtHR, 2 February 2016). For a summary of the notable cases and some less-known applications in this regard, see Päivi Korpisaari, 'Sananvapaus verkossa: yksilöön kohdistuva vihapuhe ja verkkoalustan ylläpitäjän vastuu' (2019) Lakimies 928, 948.
[596] ibid.
[597] *Ahmet Yıldırım v Turkey* ECHR 2012–VI 505, para 48; and *Times Newspapers Ltd (nos 1 and 2) v the United Kingdom* ECHR 2009–I 377, para 27.
[598] *Cengiz and Others v Turkey* ECHR 2015–VIII 177, paras 51, 52.
[599] *Egill Einarsson v Iceland* App no 24703/15 (ECtHR, 7 November 2017), para 46.
[600] *Cengiz and Others v Turkey* ECHR 2015–VIII 177, para 52.

Internet do not flow only from the content but also from facilitation,[601] the ECtHR has not so far delved deeper to scrutinize the platform mechanisms to forge a link between harm amplification and soft content moderation.

However, in other contexts, the ECtHR has reflected the role of private activity for the public interest. In *Appleby*, the ECtHR noted there could be situations where a completely privatized public sphere would bar 'any effective exercise of freedom of expression or it can be said that the essence of the right has been destroyed'. In those situations, the positive obligation of the state to 'protect the enjoyment of the Convention rights by regulating property rights' could be triggered.[602] The case concerned a private shopping center where the applicants had been prevented from expressing political views. In that case, the ECtHR noted that the applicants had had several other means and accessible places for expression. Thus, the positive obligation was not triggered.[603] However, shopping centers are, of course, spatially very limited.[604] One could claim that the scope of privately controlled platform networks is something different. This is also highlighted by the statement that YouTube lacks viable alternatives for expression.

While it is not certain what the positive obligation anchored in *Appleby* reasoning would now require, it implies that the regulation mandate of the state would contain a 'must-carry' requirement: people must be allowed to express their views on platforms.[605] Indeed, the ECtHR has often linked positive obligations of free expression to the public need to guarantee pluralism in media sector.[606] It has quite firmly stated that 'the State is the ultimate guarantor of pluralism'.[607] It's unclear, however, whether and how the pluralism argument translates to the platform context to justify the deeply needed state interference platform mechanisms. Although, the disproportionate amplification of harm with downplaying of other voices could be addressed in terms of pluralism.[608]

The ECtHR has also considered situations where certain activities of public importance have been delegated to private actors. As regards privatized schools and psychiatric treatment, the ECtHR noted 'that the State cannot absolve itself from responsibility by delegating its obligations to private bodies or individuals'.[609] In case *Sychev v Ukraine* which considered privatized execution of court judgements, it noted:

> [T]he fact that a State chooses a form of delegation in which some of its powers are exercised by another body cannot be decisive for the question of State responsibility ratione personae. In the Court's

[601] *Delfi v Estonia* ECHR 2015–II 319, paras 110, 133.
[602] *Appleby and Others v the United Kingdom* ECHR 2003–VI 165, para 47.
[603] ibid para 48.
[604] Schwarz (n 34) 122–123.
[605] On must-carry claims, see Keller (n 273) 15–27.
[606] Eg, *Centro Europa 7 Srl and Di Stefano v Italy* ECHR 2012–III 339, para 134; *Animal Defenders International v the United Kingdom* 2013–II 203, para 111.
[607] *Animal Defenders International v the United Kingdom* 2013–II 203, para 101.
[608] Helberger (n 322) 4–5.
[609] *Costello-Roberts v the United Kingdom* App no 13134/87 (ECtHR, 25 March 1993), para 27; and later *Storck v Germany* ECHR 2005–VI 111, para 103.

> view, the exercise of State powers which affects Convention rights and freedoms raises an issue of State responsibility regardless of the form in which these powers happen to be exercised.[610]

In other words, despite the delegation of State activities to private actors, at least it's still possible for the ECtHR to scrutinize the actions in the light of positive obligations. However, from case-law it becomes equally plain that the privatization of State functions is not a breach *per se*.[611] More specifically, the Court has required legislative and administrative preventative measures to be in place for maintaining adequate level of protection.[612] But the question of the adequacy of the framework in our context remains unanswered.

Finally, possible restrictions on freedom of expression through self-/co-regulatory codes of conduct and open-ended provisions, and thus the framework more generally, could be challenged based on the ECtHR's case-law on rule of law, or more precisely on the requirement of qualitative legality.[613] Under article 10(2) of the ECHR, restrictions must be 'provided by law', which requires that there is a sufficient legal basis.[614] In theory, all the regulatory instruments in our three case studies could struggle with this requirement either because there is no law or it's unprecise. In practice, it remains to be seen whether this branch of ECtHR jurisprudence will mandate changes in the context of social media.

In sum, so far it remains unclear to what extent state can outsource power and responsibility. In our context, we should also remember the challenges of scale, transnationalism and complexity. After all, a positive obligation must consider 'the choices which must be made in terms of priorities and resources' and it must not be 'interpreted in such a way as to impose an impossible or disproportionate burden on the authorities'.[615] Having said that, it appears that the ECtHR case-law imposes some restrictions on unfettered privatization. Ultimately, the exact proportion of public and private power in the regulatory framework is to be determined in the conflicts over fair balance. The doctrinal elements of the case-law I outlined may provide points of friction against the emergence of increasingly obscure and undemocratic management. Friction may sometimes result in legal conflicts before the ECtHR, which in turn might eventually produce compromises between efficiency and public accountability. While it of course ultimately depends on the European legislators to implement any additional safeguards, the ECtHR could at least push them toward the values of public accountability and openness through positive obligations. Even so, the pre-requisite conflicts before the ECtHR are yet to appear.

---

[610] *Sychev v Ukraine* App no 4773/02 (ECtHR, 11 October 2005), para 54.
[611] Stephanie Palmer, 'Privatization and Human Rights in the United Kingdom' in Tsvi Kahana and Anat Scolnicov (eds), *Boundaries of State, Boundaries of Rights* (CUP 2016) 249.
[612] *Sychev v Ukraine* App no 4773/02 (ECtHR, 11 October 2005), para 55; and *Storck v Germany* ECHR 2005–VI 111, paras 103, 107.
[613] See Chapter 3.2.2.
[614] *Editorial Board of Pravoye Delo and Shtekel v Ukraine* ECHR 2011–II 393, para 51.
[615] *Osman v the United Kingdom* ECHR 1998-VIII 101, para 116; and *Appleby and Others v the United Kingdom* ECHR 2003–VI 165, para 40.

5.2 Summary and the Threat of Co-optation

In this thesis I have traced the emerging entanglements of public and private power in freedom of expression and related rights protection on social media. I have used a piece of doctrine formulated by the ECtHR in the 60s to work as a backdrop for changes observed. The responsibility and power over fundamental rights protection has traditionally laid on the institutions of liberal democratic state. In addition, I have sought to deconstruct utilitarian efficiency as the prime driver for change by revealing and analyzing the efficiency/democracy contradiction.

Firstly, we have seen that, on the one hand, the responsibility/power is increasingly transferred to private actors, most notably to a handful of large US-based social media companies, from which it may channeled further to form complex regulatory chains. On the other hand, the responsibility/power is transferred to a suite of specialized public agencies. At the same time, regulation channels power away from democratic legislatures, including also the European Parliament, and also courts. The increasingly privatized, obscure and specialized institutional configuration distances power away from the European public of course as well.

Secondly, we have seen that the new legislative and administrative framework deploys novel regulatory techniques that engage the executive branch and private actors in cooperation to protect relevant fundamental rights on platforms. Freedom of expression protection becomes a) based on open-ended norms of binding law and different self-/co-regulatory codes of conduct; b) heavily reliant on companies' compliance apparatuses of content moderation that are tasked to carry out the 'field work' protection under the management of company experts; c) quantified to be presented in outcome-based performance reports which provide the basis for public oversight; and d) increasingly enforced through different computational algorithms that curate, rank, detect and filter information. In addition, e) individual violations are increasingly redressed through online dispute resolution mechanisms and f) systemic oversight is provided by administrative agencies vested with broad discretionary mandates with some preference for cooperative, 'bargaining' enforcement.

The framework is now leaning towards efficiency. To be sure, the structure could offer more effective ways to control user behavior to protect safe public sphere when the EU and Member States are facing the challenges of scale, transnationalism and complexity. However, we see that it also squarely fits in the stencil of neoliberal governmentality delineated by privatization, utilitarian methods for assessing costs and benefits, procedural informality, reliance on vague standards, and mediation by expert professional networks. Of course, it would be possible to turn the new framework oriented toward public accountability. We indeed noticed attempts to inject these values, for instance, in EDRi's challenge of EU Internet Forum, in the concept of organizational responsibility, and in (some of) the Parliament's legislative amendments to terrorist content regulation.

Thirdly, we have seen that the new hybrid framework sets some tension with the traditional doctrine of fundamental rights structure which understood the liberal democratic state as the guarantor of fundamental rights enjoyment and, at the same time, the interpreter of rights' meaning. Yet so far, the ECtHR jurisprudence is not yet properly developed to meet the challenges of the emerging framework. Given the challenges of scale,

transnationalism and complexity, in my view it seems likely that the Strasbourg Court will adapt the old doctrine to accommodate the increased private involvement. However, the old doctrine is a potential point of friction for structural change, thus affording opportunities to re-orient the framework toward democratic values as well.

But why, after all, does it matter that platform regulation is opaque and undemocratic, if the aim of the regulation is to protect fundamental rights, that is, if the aim is simply 'just'? To answer that we must revert to the purpose of fundamental rights structure which is fostering the political character of rights. The power-centralizing framework may squeeze the political character out of freedom of expression in digital context. This happens when democratic institutions and citizens lose their grip on interpreting what freedom of expression in the digital environment means. More specifically, the platform companies may co-opt freedom of expression on social media. Its protection turns to management. Free expression management is aligned with overall wealth-maximizing efficiency and entrenched in the political economy of social media: virtuousness of private innovation, individual user empowerment, and disproportionate anxiety over public interference. Then freedom of expression falls to irrelevance – to a card like the US First Amendment that could be always used by the companies to entrench their information sovereignty.

Yet co-optation makes it harder to address some very real substantive freedom of expression issues too. It becomes considerably harder to claim that private innovation may endanger and not only benefit free expression; that free expression should also protect against new systemic harms such as wide-scale manipulation and reduction of social cohesion; that freedom of information cannot be reduced to a 'more is better' maxim; and that threats to free expression may in fact flow directly from the self-serving interests of private companies which purport to advance the very same right. It should be explicitly recognized that while the platforms have advanced people's free expression in many ways that are well-known, they can also be damaging to not only people's privacy and other rights, but also to their freedom of expression. It is a twisted condition of neoliberal age that freedom of expression issues are perhaps more politicized than in a long time but, at the same time, especially small Member States have lost the power to set the overall terms for democratic deliberation that is now mediated by these platforms. Small states' power is limited to putting *ex post* plasters by, for instance, chasing individual perpetrators in collective campaigns of hate and harassment while the defunct online systems manipulate their citizens and erode their democracies.

5.3 Toward Compromise? – Final Remarks

While my thesis has critiqued the role of social media companies in regulation from the perspective of the positive obligations, I don't propose that we should simply go back to the old. Neither I'm suggesting that European regulators should 'take over' the activity of content moderation from companies entirely, i.e. that private companies should not have positive obligations at all, if that is even possible given the slippery line of positive/negative obligations and the interconnectedness of expression and a medium. Also, we should remember that the regulation in case studies, despite their numerous managerial traps, contained also potential

for political equality. Even so, I haven't hidden my proposition that the overseeing role of democratic institutions in regulation should be significantly stronger to avoid the co-optation of free expression. Thus, before wrapping up, I want to tentatively outline some reconstructions for a better compromise between efficiency and democracy. I believe these suggestions may also serve to preserve (the old) rule of law to which fundamental rights are inextricably linked and, thus, prevent constitutional rot more generally.

So, firstly I believe I have not left unclear that freedom of expression in the social media age needs reformulation. When the idea of rational individual collapses, and there exists far more insidious ways to make people's voice disappear and democracies crumble than just archaic censorship, also the right itself should transform to accommodate new harms. Information manipulation should be considered a threat for freedom of expression and information. Society's eroding ability to come together to deliberate and compromise on the hard questions of good and bad should be considered a threat for freedom of expression and information. Algorithmic discrimination and amplification of harm on platforms should be considered a threat for freedom of expression and information. Perhaps these issues are approachable from the perspective of other rights. However, it is not that other rights would necessarily be less prone to co-optation than free expression, which in my view still has special normative force in Western societies. Thus, it's a worthy site for contesting power.

However, 'the realization of rights is always contingent on enabling institutions'.[616] Thus, secondly, with the emerging framework that entangles public and private in intricate ways, it is necessary to greatly enhance the horizontality of fundamental rights by envisioning new ways for people and communities to assert their rights against unprecedented private power. As Cohen states, 'institutions for recognizing and enforcing fundamental rights should work to counterbalance private economic power rather than reinforcing it. Obligations to protect fundamental rights must extend – enforceably – to private, for-profit entities if they are to be effective at all'.[617] In particular, the doctrine of horizontal effect should enable cracking open the proprietary and opaque systems of algorithmic intermediation. It is clear, as we have seen in this thesis, that private power is facilitated by law, which in a hierarchical legal system is ultimately traced back to a constitution. Trade secrecy and IPR would not exist constitutional law, that is, without the right to conduct business, right to property, law-enforcing institutions and so on. So, mending the undue power imbalances must also be done by constitutional law. This is especially true when under neoliberal governmentality the state itself is turning to an image of a company, animated by the market logic of efficiency.

My thesis has focused on the issue of state-mediated horizontality. However, also indirect and direct horizontality should be reinforced, an effort that would more precisely focus on the protective potential of courts. Yet the realities of information age, to some extent, seem to demand for a shift of fundamental rights oversight from courts to more specialized administrative entities (or at least to more specialized courts). But in the long run, an institutional reform to the court system should be made so that the relevance of courts in

[616] Cafaggi and Pistor (n 72)
[617] Cohen (n 17) 267.

adjudicating the meaning of rights in asymmetrical horizontal relations could be upheld. For instance, horizontality should be more firmly linked with the non-domination principle. By focusing on potential power abuse, one could interrogate private power that eludes individual awareness.

Secondly, the EU and Member States should look for new ways to put their increasingly powerful administrative agencies under democratic scrutiny as well. Here, I am inspired by Laidlaw, who suggests not only an agency for overseeing social media and other gatekeepers from the fundamental rights perspective, but also a parliamentary committee that would oversee the operations of this agency.[618] In the information age characterized by increasing complexity, a workable compromise for democratic accountability may require certain specialization from legislators as well. It should be noted that oversight might lead to increased localization of platforms too. Yet a pre-condition to the empowerment of Member State institutions is that the EU not only channels power away from platform companies, but also that it re-distributes it sufficiently evenly. While the concerns over populist authoritarianism are indeed warranted, and EU-level (and even global) coordination is obviously needed as well, it is not a fair compromise to centralize all the power to the Commission, EU agencies or Ireland either. The undeniable challenge here is that such an exercise would require the EU to pay less attention to its internal market roots that underscore efficiency.

Since I noted that positive obligations are mandates directed also to legislators, I regard my thesis a timely one. As the EU lumbers forward with several major initiatives, including the Digital Services Act that is posited to refresh social media's responsibility for content,[619] one can expect conflict-ridden years ahead. Companies' lobbying is well under way with Google, Facebook and EDiMA spearheading the efforts to spread the presumptive virtuousness of private ordering, algorithmic intermediation, and unencumbered innovation. Elsewhere, within the European Parliament there seems to be an explicit intention to have the platform businesses scrutinized.[620] And of course, the CJEU has not become entirely disempowered either.[621] While much is at this point uncertain, competition over the sovereignty of information flows is bound to continue. What remains to be seen is whether these conflicts will end up in a compromise for the benefit of European people.

[618] Laidlaw (n 15) 251.
[619] European Parliament, 'Draft Report with recommendations to the Commission on Digital Services Act: Improving the functioning of the Single Market' 2020/2018 (INL).
[620] Eg, European Parliament, 'Report on competition policy – annual report 2019' P9_TA (2020) 0158, para 105, where the Parliament 'calls on the Commission to ban platforms from displaying micro-targeted advertisements'.
[621] See Case C-311/18 *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems*, EU:C:2020:559.